

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:17:38 ; Search time 21.3103 Seconds
(without alignments)
58.013 Million cell updates/sec

Title: US-09-496-391-4
Perfect score: 27
Sequence: 1 AKKARA 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SPREMBL_21.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_phage.*
- 11: sp_plant.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriophage.*
- 17: sp_archaeop.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	27	100.0	156	10	09S192
2	27	100.0	317	16	08X85
3	27	100.0	366	10	08S1V6
4	27	100.0	368	5	09XYS5
5	27	100.0	411	16	0983E6
6	27	100.0	749	17	09YFA1
7	27	100.0	987	4	09BUT7
8	27	100.0	998	4	09P211
9	24	88.9	22	13	09P865
10	24	88.9	70	16	09PBR7
11	24	88.9	79	2	09X9G0
12	24	88.9	100	16	08X176
13	24	88.9	100	17	09YB2
14	24	88.9	119	5	P90596
15	24	88.9	135	5	09U701
16	24	88.9	135	5	09U917

17	24	88.9	135	5	002617	002617 trypanosoma
18	24	88.9	135	5	027794	027794 trypanosoma
19	24	88.9	135	5	08T2X9	08T2X9 trypanosoma
20	24	88.9	139	11	09CSG5	09CSG5 mus musculus
21	24	88.9	156	16	P94450	P94450 bacillus su
22	24	88.9	156	16	0927G8	0927G8 listeria in
23	24	88.9	156	16	08Y405	08Y405 listeria mo
24	24	88.9	166	5	099281	099281 paracanthus
25	24	88.9	170	16	08ZPV3	08ZPV3 yersinia pe
26	24	88.9	178	5	044698	044698 caenorhabdi
27	24	88.9	180	5	025636	025636 paracanthus
28	24	88.9	186	10	09S9S7	09S9S7 arabidopsis
29	24	88.9	190	10	09ASC7	09ASC7 oryza sativ
30	24	88.9	192	15	091131	091131 human immun
31	24	88.9	195	12	081836	081836 hepatitis d
32	24	88.9	211	16	098410	098410 rhizobium l
33	24	88.9	214	12	081835	081835 hepatitis d
34	24	88.9	227	5	023792	023792 chironomus
35	24	88.9	250	16	093ME6	093ME6 clostridium
36	24	88.9	262	2	08RTS3	08RTS3 uncultured
37	24	88.9	269	3	096M03	096M03 candida mal
38	24	88.9	269	16	08U919	08U919 agrobacteri
39	24	88.9	278	4	09NTB5	09NTB5 homo sapien
40	24	88.9	286	4	09P0C2	09P0C2 homo sapien
41	24	88.9	301	11	09D8C0	09D8C0 mus musculi
42	24	88.9	302	10	09FM28	09FM28 arabidopsis
43	24	88.9	310	2	09AUS9	09AUS9 shigella fl
44	24	88.9	312	10	09PT25	09PT25 phaseolus v
45	24	88.9	315	16	08Y107	08Y107 raietonia s

ALIGNMENTS

RESULT 1

ID 09S192 PRELIMINARY: PRT: 156 AA.

AC 09S192

DT 01-MAY-2000 (TREMUREL, 13, Created)

DT 01-MAY-2000 (TREMUREL, 13, Last sequence update)

DT 01-DEC-2001 (TREMUREL, 19, Last annotation update)

DE At2g11830 protein.

GN At2g11830.

GN Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:

OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:

OC eucosids II: Brassicales: Brassicaceae: Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RX MEDLINE=20083487; PubMed=10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., Vanden S.E., Umayam L., Tallon L.J., Gill J.E.,

RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;

RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

thaliana.";

RL Nature 402:761-768(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC007187; AAD28677.1; -

SO SEQUENCE 156 AA; 18095 MW; 4230A87CE3BPF812 CRC64;

Query Match 100.0%; Score 27; DB 10; Length 156;

Best Local Similarity 100.0%; Pred. No. 80;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps

OY 1 AKKARA 6
111111
DB 144 AKKARA 149

RESULT 2

08XX85 PRELIMINARY; PRT; 317 AA.
AC 08XX85;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Hypothetical protein RSC2231.
GN RSC2231 OR RS01368.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM100;
RX MEDLINE=21681879; PubMed=11823852;
RA Salenouat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Arlet M., Billault A., Brotier P., Camus J.C., Cactolico L., Chandel M., Choise N., Claudel-Renard C., Cunac S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T., Sigler P., Thebault P., Whalen M., Wincker P., Levy M., Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 413:497-502(2002).
DR EMBL; AL646069; CAD15938.1; -
DR InterPro; IPR005269; Cons_hypoth730.
DR Pfam; PF03641; Lysine decarbox. 1.
DR TIGRfams; TIGR00730; TIGR00730.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 317 AA; 34479 MW; E221A9B642011507 CRC64;

Query Match 100.0%; Score 27; DB 16; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
111111
DB 94 AKKARA 99

RESULT 3
08S5V6 PRELIMINARY; PRT; 366 AA.
AC 08S5V6;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Putative NAM-like protein.
GN OJ1015F07.11.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=3947;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T., Sasaki C., Henry D., Oates R., Simmons J.;
RT "Rice Genomic Sequence.";
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC104427; AAM1913.1; -
SQ SEQUENCE 366 AA; 40400 MW; C87022A614E1D53E CRC64;

Query Match 100.0%; Score 27; DB 10; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
111111
DB 265 AKKARA 270

RESULT 4

09XX55 PRELIMINARY; PRT; 368 AA.
AC 09XX55;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE Dihydroliipoyl dehydrogenase-binding protein.
OS Ascaris suum (pig roundworm) (Ascaris lumbricoides).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92296744; PubMed=1376597;
RA Komuniecki R., Rhee R., Bhat D., Duran E., Sidawy E., Song H.;
RT "The pyruvate dehydrogenase complex from the parasitic nematode Ascaris suum: novel subunit composition and domain structure of the dihydroliipoyl transacylase component.";
RL Arch. Biochem. Biophys. 296:115-121(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96214998; PubMed=8621401;
RA Klingbeil M.M., Walker D.J., Arnette R., Sidawy E., Hayton K., Komuniecki P.R., Komuniecki R.;
RT "Identification of a novel dihydroliipoyl dehydrogenase-binding protein in the pyruvate dehydrogenase complex of the anaerobic parasitic nematode, Ascaris suum.";
RL J. Biol. Chem. 271:5451-5457(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Arnette R.L., Komuniecki R.;
RL Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF12459; AAD30034.1; -
DR HSSP; P07016; 1CA4F.
DR InterPro; IPR001078; 2oxoacid_dh.
DR InterPro; IPR004167; E3_binding.
DR Pfam; PF00198; 2-oxoacid_dh; 1.
DR Pfam; PF02817; e3_binding; 1.
DR Prodom; PD001115; 2oxoacid_dh; 1.
SQ SEQUENCE 368 AA; 39089 MW; F49876B085622DD4 CRC64;

Query Match 100.0%; Score 27; DB 5; Length 368;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
111111
DB 258 AKKARA 263

RESULT 5

0983E6 PRELIMINARY; PRT; 411 AA.
AC 0983E6;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE Hypothetical protein mlr8361.
GN Mlr8361.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;

Query Match 100.0%; Score 27; DB 5; Length 368;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003013; BAB53935.1; -
 DR InterPro: IPR002549; UPF0118.1.
 DR Pfam: PF01594; UPF0118.1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 411 AA; 44013 MW; 27D23F289B3AA31D CRC64;

Query Match 100.0%; Score 27; DB 16; Length 411;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
 Db 406 AKKARA 411

RESULT 6
 Q9YFA1 PRELIMINARY; PRT; 749 AA.

AC Q9YFA1:
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein APE0340.
 GN APE0340.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
 OC Desulfurococcaceae; Aeropyrum.
 NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1.
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Nakai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000059; BAA79295.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 749 AA; 85113 MW; 8140779388D37ADD CRC64;

Query Match 100.0%; Score 27; DB 17; Length 749;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
 Db 48 AKKARA 53

RESULT 7
 Q9BUT7 PRELIMINARY; PRT; 987 AA.

AC Q9BUT7:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 110.2 KDa protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC002575; AAH02575.1; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00271; Helicase_C.1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELICC; 1.
 KW Hypothetical protein.
 FT NON-TER 1
 SQ SEQUENCE 987 AA; 110202 MW; 590EF94358750924 CRC64;

Query Match 100.0%; Score 27; DB 4; Length 987;
 Best Local Similarity 100.0%; Pred. No. 4.9e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
 Db 354 AKKARA 359

RESULT 8
 Q9P211 PRELIMINARY; PRT; 998 AA.

AC Q9P211:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE KIA1517 protein (Fragment).
 GN KIA1517.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20277482; PubMed=10819331;
 RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human
 RT genes. XVII. The complete sequences of 100 new cDNA clones from brain
 RT which code for large proteins in vitro.";
 RL DNA Res. 7:143-150(2000).
 DR EMBL: AB040950; BAA9604.1; -
 DR InterPro: IPR002106; AALRNA_ligaseII.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00271; Helicase_C.1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELICC; 1.
 DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
 FT NON-TER 1
 SQ SEQUENCE 998 AA; 110707 MW; 5B330476537C8944 CRC64;

Query Match 100.0%; Score 27; DB 4; Length 998;
 Best Local Similarity 100.0%; Pred. No. 5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
 Db 347 AKKARA 352

RESULT 9

AC Q9PS65 PRELIMINARY; PRT; 22 AA.
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

```

DE 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
OS Histone H1B (Fragment).
OC Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP MEDLINE=92082492; PubMed=1747124;
RA Davie J.R., Delcuve G.P.;
RT "Characterization and chromatin distribution of the H1 histones and
RT high-mobility-group non-histone chromosomal proteins of trout liver
RT and hepatocellular carcinoma."
RL Biochem. J. 280:491-497(1991).
FT NON_TER 1
FT 22
FT 22
SQ SEQUENCE 22 AA; 2132 MW; 3E90388F68189AE3 CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 13; Length 22;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
Db 11 AKKARA 16

RESULT 10
O9PBR7 PRELIMINARY; PRT: 70 AA.
AC O9PBR7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE Hypothetical protein Xf2073.
GN Xf2073.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RA MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franco S.C., Franco M.C., Frohne M., Furlan L.R.,
RA Gardier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krueger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madalena A.M.B.N., Madeira H.M.F., Marino C.Y.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miranca E.C., Miyaki C.Y., Monteleiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quadrio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Trufi D., Tsai S.M., Tsuchiko M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RT Nature 406:151-155(2000).

```

```

DR EMBL: AE004023; AAF84872.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 70 AA; 8084 MW; A8BDC8ACD606E951 CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 16; Length 70;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
Db 22 AKKARA 27

RESULT 11
O9X9G0 PRELIMINARY; PRT: 79 AA.
AC O9X9G0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE Hypothetical 9.2 kDa protein.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99426823; PubMed=10496905;
RA Rakin A., Noelling C., Schubert S., Heesemann J.;
RT "Common and specific characteristics of the high-pathogenicity island
RT of Yersinia enterocolitica."
RL Infect. Immun. 67:5265-5274(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Rakin A., Schubert S., Gullvout I., Carniel E., Heesemann J.;
RT "Local hopping of IS3 elements into the A+T-rich part of the high-
RT pathogenicity island in Yersinia enterocolitica 1B, O:8."
RL FEMS Microbiol. Lett. 182:2255-2294(2000).
DR EMBL: AJ236887; CAB46602.1; -.
KW Hypothetical protein.
SQ SEQUENCE 79 AA; 9199 MW; 62F484D339D5F45 CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 2; Length 79;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
Db 28 AKKARA 33

RESULT 12
O8X176 PRELIMINARY; PRT: 100 AA.
AC O8X176;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE Hypothetical protein CPE2245.
GN CPE2245.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 // TYPE A;
RA PubMed=11792842;
RA Shmizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh eater."
RT Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

```


DR EMBL; AP003193; BAB81951.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 100 AA; 11749 MW; 42FC03397837AFB0 CRC64;
Query Match 88.9%; Score 24; DB 16; Length 100;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 AKKARA 6
Db 56 AKKARA 61
RESULT 13
O9YBD2
ID O9YBD2; PRELIMINARY; PRT; 100 AA.
AC O9YBD2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein APE1665.
GN APE1665.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
ER EMBL; AP000062; BAA80666.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 100 AA; 10715 MW; 9C7F1F634802104A CRC64;
Query Match 88.9%; Score 24; DB 17; Length 100;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 AKKARA 6
Db 41 SKKARA 46
RESULT 14
P90596
ID P90596; PRELIMINARY; PRT; 119 AA.
AC P90596;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Histone H2A (Fragment).
OS Trypanosoma cruzi.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL-BRENER;
RX MEDLINE=20018171; PubMed=10548721;
RA Vazquez M., Lorenzi H., Schlijman A., Ben-Dov C., Levin M.;
RT "Analysis of the distribution of SIRE in the nuclear genome of
RT Trypanosoma cruzi.";
RL Gene 239:207-216(1999).
ER EMBL; Y10371; CAA71400.1; -
DR InterPro; IPR004822; Histone_core.

DR InterPro; IPR002119; Histone_H2A.
DR Pfam; PF00125; histone.1.
DR PRINTS; PR00620; HISTONEH2A.
DR PRODOM; PD000522; histone_H2A; 1.
DR SMART; SM00414; H2A; 1.
DR PROSITE; PS00046; HISTONE_H2A; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 119 AA; 12654 MW; A51E785CB47CE6F CRC64;
Query Match 88.9%; Score 24; DB 5; Length 119;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 AKKARA 6
Db 110 SKKARA 115
RESULT 15
O9U701
ID O9U701; PRELIMINARY; PRT; 135 AA.
AC O9U701;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Histone H2A unit A.
GN H2A.
OS Trypanosoma cruzi.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=20175181; PubMed=10708660;
RA Thomas M.C., Olivares M., Escalante M., Maranon C., Montilla M.,
RA Nicholas S., Lopez M.C., Puerta C.;
RT "Plasticity of the histone H2A genes in a Brazilian and six Colombian
RT strains of Trypanosoma cruzi.";
RL Acta Trop. 75:203-210(2000).
ER EMBL; AF167074; AAD54272.1; -
DR InterPro; IPR004822; Histone_core.
DR InterPro; IPR002119; Histone_H2A.
DR Pfam; PF00125; histone.1.
DR PRINTS; PR00620; HISTONEH2A.
DR PRODOM; PD000522; histone_H2A; 1.
DR SMART; SM00414; H2A; 1.
DR PROSITE; PS00046; HISTONE_H2A; UNKNOWN_1.
SQ SEQUENCE 135 AA; 14336 MW; DC41BC38E7CCDDB CRC64;
Query Match 88.9%; Score 24; DB 5; Length 135;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 AKKARA 6
Db 126 SKKARA 131

Search completed: April 8, 2003, 11:28:11
Job time : 25.3103 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:11:45 ; Search time 33.7931 Seconds
(without alignments)
31.545 Million cell updates/sec

Title: US-09-496-391-5

Perfect score: 36

Sequence: 1 ARKRAKA 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	8	21	Peptide modulating
2	36	100.0	14	17	AKAP79 AA38, 39 mu
3	36	100.0	14	21	Mutant peptide AKA
4	36	100.0	22	21	Peptide modulating
5	36	100.0	24	21	Peptide modulating
6	36	100.0	24	21	Peptide modulating
7	36	100.0	32	21	Peptide modulating
8	36	100.0	32	21	Peptide modulating
9	36	100.0	40	21	Peptide modulating
10	33	91.7	8	21	Peptide modulating

11	33	91.7	23	21	Peptide modulating
12	33	91.7	31	21	Peptide modulating
13	32	88.9	14	17	AKAP79 (residues 3
14	32	88.9	14	21	Human AKAP 79 pept
15	32	88.9	331	21	Arabidopsis thalia
16	32	88.9	331	21	Arabidopsis thalia
17	32	88.9	471	22	Arabidopsis thalia
18	32	88.9	478	21	Arabidopsis thalia
19	32	88.9	478	21	Arabidopsis thalia
20	32	88.9	564	22	Arabidopsis thalia
21	32	88.9	2059	23	Streptococcus poly
22	31	86.1	24	21	Peptide modulating
23	30	83.3	8	21	Peptide modulating
24	30	83.3	14	21	Antimicrobial pro-
25	30	83.3	14	22	Synthetic anti-mic
26	30	83.3	14	23	Pro-apoptosis pept
27	30	83.3	14	23	Selective targetin
28	30	83.3	14	23	Pro-apoptosis pept
29	30	83.3	16	21	Peptide modulating
30	30	83.3	16	17	Amphiphilic antimi
31	30	83.3	18	22	18 mer peptide seq
32	30	83.3	21	22	Novel human diago
33	30	83.3	159	22	Novel human diago
34	29	80.6	10	22	Human complementar
35	29	80.6	43	22	Peptide #10837 enc
36	29	80.6	43	22	Human brain expres
37	29	80.6	43	22	Peptide #11217 enc
38	29	80.6	151	22	Novel human diago
39	29	80.6	160	18	Streptococcus pneu
40	29	80.6	227	20	Neisseria meningit
41	29	80.6	227	23	Neisseria meningit
42	29	80.6	360	21	Human interferon-a
43	29	80.6	361	22	Novel human diago
44	29	80.6	420	22	Novel human diago
45	29	80.6	428	22	Novel human diago

ALIGNMENTS

RESULT 1

AA08137

ID AAB08137 standard; peptide: 8 AA.

AC AAB08137;

DT 04-DEC-2000 (first entry)

XX Peptide modulating activity of heparin, and other glycans.

DE Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;

XX cell attachment; cell adhesion; vein graft; tumour cell metastasis;

KW cartilage differentiation; wound healing.

XX Synthetic.

OS Key

XX Misc-difference 1..8

FT location/Qualifiers

FT note="this peptide may be repeated an

FT unspecified number of times"

PN WO20045831-A1.

XX 10-AUG-2000.

PD 02-FEB-2000; 2000WO-US02853.

PF 02-FEB-1999; 99US-0118276.

PR (UJBE-) UNIV JEFFERSON THOMAS.

XX San Antonio JD, Verrecchio A, Schlick BP;

DR WPI: 2000-543446/49.

XX

PT Novel synthetic peptides with high affinity for glycoaminoglycans and

PT proteoglycans, useful for modulating heparin, promoting cell

PT attachment, modulating tumour metastasis and modulating wound healing -

XX

PS Disclosure; Page 23; 76pp: English.

XX

CC The present sequence represents a synthetic peptide which has a high

CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful

CC in methods for modulating heparin or other glycoaminoglycans with

CC anticoagulant activity, promoting cell attachment or adhesion to

CC natural or synthetic surfaces (especially vein grafts), modulating

CC tumour cell metastasis, modulating cartilage differentiation, targeting

CC drugs to epithelial cell surfaces (or to other cells expressing

CC proteoglycans), modulating enzymes that act on glycoaminoglycan

CC substrates, affinity purification of bioactive sequences of a

CC glycoaminoglycan, modifying endothelial cell pro-coagulant or

CC anti-coagulant functions mediated through glycoaminoglycans, and

CC modulating wound healing. The peptide may also be used for blocking

CC tissue uptake of heparin or other glycoaminoglycans in a mammal to

CC increase heparin half-life in circulation.

CC

XX

SQ Sequence 8 AA:

Query Match 100.0%; Score 36; DB 21; Length 8;

Best Local Similarity 100.0%; Pred. No. 7.8e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAAGA 8

Db 1 ARKKAAGA 8

RESULT 2

AAW02565

ID AAW02565 standard; peptide; 14 AA.

XX

AC AAW02565;

XX

DT 13-JAN-1997 (first entry)

XX

DE AKAP79 AA38, 39 mutation.

XX

KW Autoimmune disease; PACT59; PACT74; PACT36; PACT60; murine; T-cell; PKA;

KW mouse; A-kinase anchoring protein 79; CAMP-dependent protein kinase;

KW postsynaptic density; AKAP79; human forebrain; transcriptional activator;

KW interleukin 2; T-cell activation; calcineurin; T-cell clonal expansion;

KW calcium/calmodulin dependent protein phosphatase; T-cell response;

KW autoimmunity related disease; therapy; immune response.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 3 /note="Arg40Ala"

FT

XX

PN W09616172-A2.

XX

PD 30-MAY-1996.

XX

PF 22-NOV-1995; 95MO-US16039.

XX

XX

PR 17-JUL-1995; 95US-0503226.

PR 23-NOV-1994; 94US-0344227.

PR 15-MAR-1995; 95US-0404731.

XX

PA (ICOS-) ICOS CORP.

PA (ORBS-) STATE OF OREGON.

PA (UYOR-) UNIV OREGON HEALTH SCI.

XX

PI Cognahan VM, Gallatin WM, Howard ML, Lockerbie RO;

PI Scott JD;

XX

DR WPI: 1996-268608/27.

XX

PT New modulators of anchoring protein function - used to develop

PT prods. for use in the treatment of auto-immune-related conditions.

XX

PS Example 11; Page 34; 74pp: English.

XX

CC AAW02565-W02569 represent mutants of residues 37-50 of the A-kinase

CC anchoring protein 79 (AKAP79). AKAP79 (see AAW05264 for wild type

CC residues 37-50) is responsible for anchoring CAMP-dependent protein kinase

CC (PKA) to specific intracellular sites. AKAP79 is predominantly present

CC in postsynaptic densities in the human forebrain. The pathways that

CC involve AKAP79 are important in many cell types and have been implicated

CC in many cell functions, including the transcriptional activation of the

CC interleukin 2 gene that is important in T-cell activation. AKAP also

CC binds to calcineurin (see AAW02536), which is a calcium/calmodulin

CC dependent protein phosphatase associated with T-cell activation. By

CC binding both PKA and calcineurin, AKAP79 co-localises a kinase and a

CC phosphatase which may regulate flux through a specific signalling

CC pathway. The AKAP79 binding sequences can be used to develop products

CC for use in the treatment of autoimmune related conditions. The AKAP79

CC binding proteins can be used in methods for stimulating an immune

CC response, and for stimulating activated T-cells for selected clonal

CC expansion. The proteins can also be used in a method for enhancing

CC T-cell responses to experimental stimuli for evaluation of early events

CC in T-cell biology and activation of the immune response.

CC

XX

SQ Sequence 14 AA:

Query Match 100.0%; Score 36; DB 17; Length 14;

Best Local Similarity 100.0%; Pred. No. 1.4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAAGA 8

Db 3 ARKKAAGA 10

RESULT 3

AAB14905

ID AAB14905 standard; peptide; 14 AA.

XX

AC AAB14905;

XX

DT 08-JAN-2001 (first entry)

XX

DE Mutant peptide AKAP 79 AA38, 39.

XX

KW Human; calcineurin-binding peptide; A-kinase anchor protein; AKAP;

KW AKAP 79; immunostimulant; interleukin 2 expression modulation;

KW graft rejection; transplantation; T cell-mediated disorder; mutant.

XX

OS Homo sapiens.

XX

FH Synthetic.

FT

XX

PN US6107104-A.

XX

PD 22-AUG-2000.

XX

PF 27-SEP-1996; 96US-0721458.

XX

XX

PR 23-NOV-1994; 94US-0344227.

PR 15-MAR-1995; 95US-0404731.

PR 17-JUL-1995; 95US-0503226.

XX

PA (ICOS-) ICOS CORP.

XX

PI Lockerbie RO, Gallatin WM, Lai Y, Howard ML;

XX

DR WPI: 2000-578541/54.

XX

PT Novel calcineurin deletion mutant having calcineurin polypeptide

PT sequence and binding A-kinase anchor proteins, for treating graft
PT rejection following organ transplantation and T cell-mediated disorders

PS Example 11: Column 21; 53pp; English.

XX The present sequence is a mutant peptide derived from A-kinase anchor
CC protein 79 (AKAP 79). It is expressed as a poly-histidine tag fusion
CC protein and can thus be purified to homogeneity by nickel affinity
CC chromatography. AKAP 79 binds both cAMP-dependent protein kinase (PKA)
CC and calcineurin and so co-localises a kinase and a phosphatase that
CC may regulate flux through a specific signalling pathway. Calcineurin is a
CC Ca²⁺/calmodulin-dependent protein phosphatase which is involved in many
CC intracellular signalling pathways. It participates in regulation of IL-2
CC expression following T cell stimulation in T cells. Calcineurin-binding
CC peptides derived from AKAP 79 may be used to inhibit calcineurin activity
CC in a cell. The peptides are useful for treating graft rejection following
CC organ transplantation and for treating T cell-mediated disorders.
CC Calcineurin deletion mutants which bind AKAP 79 are useful for defining
CC an AKAP 79 binding site, for stimulating the immune response, stimulating
CC activated T cells for selected clonal expansion, or for enhancing T cell
CC responses to experimental stimuli for evaluation of early events in
CC T cell biology and activation of the immune response.

XX Sequence 14 AA;

Query Match 100.0%; Score 36; DB 21; Length 14;
Best Local Similarity 100.0%; Pred. NO. 1.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAAGA 8
| | | | | | | |
DB 3 ARKKAAGA 10

RESULT 4
AAB08176
ID AAB08176 standard; peptide: 22 AA.

XX AAB08176;

DT 04-DEC-2000 (first entry)

DE Peptide modulating activity of heparin, and other glycans.

XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW cartilage differentiation; wound healing.

OS Synthetic.

XX WO200045831-A1.

XX 10-AUG-2000.

XX 02-FEB-2000; 2000WO-US02853.

XX 02-FEB-1999; 99US-0118276.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX San Antonio JD, Verrecchio A, Schlick BP;

XX WPI: 2000-543446/49.

XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -

XX Disclosure: Page 31; 76pp; English.

XX The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful

CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.

XX Sequence 22 AA;

Query Match 100.0%; Score 36; DB 21; Length 22;
Best Local Similarity 100.0%; Pred. NO. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAAGA 8
| | | | | | | |
DB 1 ARKKAAGA 8

RESULT 5
AAB08171
ID AAB08171 standard; peptide: 24 AA.

XX AAB08171;

DT 04-DEC-2000 (first entry)

DE Peptide modulating activity of heparin, and other glycans.

XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW cartilage differentiation; wound healing.

OS Synthetic.

XX WO200045831-A1.

XX 10-AUG-2000.

XX 02-FEB-2000; 2000WO-US02853.

XX 02-FEB-1999; 99US-0118276.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX San Antonio JD, Verrecchio A, Schlick BP;

XX WPI: 2000-543446/49.

XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX Disclosure: Page 30; 76pp; English.

XX The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to

CC Increase heparin half-life in circulation.

XX Sequence 24 AA;

Query Match 100.0%; Score 36; DB 21; Length 24;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAKA 8

DB 1 ARKKAKA 8

RESULT 6

AAB08178

ID AAB08178 standard; peptide: 24 AA.

XX AAB08178;

DT 04-DEC-2000 (first entry)

XX Peptide modulating activity of heparin, and other glycans.

XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;

KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;

XX cartilage differentiation; wound healing.

OS Synthetic.

PN WO200045831-A1.

PD 10-AUG-2000.

PF 02-FEB-2000; 2000WO-US02853.

PR 02-FEB-1999; 99US-0118276.

XX (UYJE-) UNIV JEFFERSON THOMAS.

PA San Antonio JD, Verrecchio A, Schick BP;

PI WPI: 2000-543446/49.

XX Novel synthetic peptides with high affinity for glycoaminoglycans and

PT proteoglycans, useful for modulating heparin, promoting cell

PR attachment, modulating tumour metastasis and modulating wound healing -

PS Disclosure: Page 31; 76pp; English.

CC The present sequence represents a synthetic peptide which has a high

CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful

CC in methods for modulating heparin or other glycoaminoglycans with

CC anticoagulant activity, promoting cell attachment or adhesion to

CC natural or synthetic surfaces (especially vein grafts), modulating

CC tumour cell metastasis, modulating cartilage differentiation, targeting

CC drugs to epithelial cell surfaces (or to other cells expressing

CC proteoglycans), modulating enzymes that act on glycoaminoglycan

CC substrates, affinity purification of bioactive sequences of a

CC glycoaminoglycan, modifying endothelial cell pro-coagulant or

CC anti-coagulant functions mediated through glycoaminoglycans, and

CC modulating wound healing. The peptide may also be used for blocking

CC tissue uptake of heparin or other glycoaminoglycans in a mammal to

CC increase heparin half-life in circulation.

XX Sequence 24 AA;

Query Match 100.0%; Score 36; DB 21; Length 24;

Best Local Similarity 100.0%; Pred. No. 2.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAKA 8

DB 1 ARKKAKA 8

RESULT 7
AAB08168
ID AAB08168 standard; peptide: 32 AA.

XX AAB08168;

DT 04-DEC-2000 (first entry)

XX Peptide modulating activity of heparin, and other glycans.

XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;

KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;

XX cartilage differentiation; wound healing.

OS Synthetic.

PN WO200045831-A1.

PD 10-AUG-2000.

PF 02-FEB-2000; 2000WO-US02853.

PR 02-FEB-1999; 99US-0118276.

XX (UYJE-) UNIV JEFFERSON THOMAS.

PA San Antonio JD, Verrecchio A, Schick BP;

PI WPI: 2000-543446/49.

XX Novel synthetic peptides with high affinity for glycoaminoglycans and

PT proteoglycans, useful for modulating heparin, promoting cell

PR attachment, modulating tumour metastasis and modulating wound healing -

PS Disclosure: Page 30; 76pp; English.

CC The present sequence represents a synthetic peptide which has a high

CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful

CC in methods for modulating heparin or other glycoaminoglycans with

CC anticoagulant activity, promoting cell attachment or adhesion to

CC natural or synthetic surfaces (especially vein grafts), modulating

CC tumour cell metastasis, modulating cartilage differentiation, targeting

CC drugs to epithelial cell surfaces (or to other cells expressing

CC proteoglycans), modulating enzymes that act on glycoaminoglycan

CC substrates, affinity purification of bioactive sequences of a

CC glycoaminoglycan, modifying endothelial cell pro-coagulant or

CC anti-coagulant functions mediated through glycoaminoglycans, and

CC modulating wound healing. The peptide may also be used for blocking

CC tissue uptake of heparin or other glycoaminoglycans in a mammal to

CC increase heparin half-life in circulation.

XX Sequence 32 AA;

Query Match 100.0%; Score 36; DB 21; Length 32;

Best Local Similarity 100.0%; Pred. No. 2.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAKA 8

DB 1 ARKKAKA 8

RESULT 8

AAB08175

ID AAB08175 standard; peptide: 32 AA.

XX AAB08175;

DT 04-DEC-2000 (first entry)

XX Peptide modulating activity of heparin, and other glycans.

XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
 KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
 KW cartilage differentiation; wound healing.
 XX Synthetic.
 OS WO200045831-A1.
 XX
 XX 10-AUG-2000.
 XX
 XX 02-FEB-2000; 2000WO-US02853.
 XX
 XX 02-FEB-1999; 99US-0118276.
 XX
 XX (UYJE-) UNIV JEFFERSON THOMAS.
 XX
 XX San Antonio JD, Verrecchio A, Schick BP;
 XX
 XX WPI: 2000-543446/49.
 XX
 XX Novel synthetic peptides with high affinity for glycoaminoglycans and
 PT proteoglycans, useful for modulating heparin, promoting cell
 PT attachment, modulating tumour metastasis and modulating wound healing -
 PS Disclosure: Page 31; 76pp; English.
 XX
 XX The present sequence represents a synthetic peptide which has a high
 CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
 CC in methods for modulating heparin or other glycoaminoglycans with
 CC anticoagulant activity, promoting cell attachment or adhesion to
 CC natural or synthetic surfaces (especially vein grafts), modulating
 CC tumour cell metastasis, modulating cartilage differentiation, targeting
 CC drugs to epithelial cell surfaces (or to other cells expressing
 CC proteoglycans), modulating enzymes that act on glycoaminoglycan
 CC substrates, affinity purification of bioactive sequences of a
 CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
 CC anti-coagulant functions mediated through glycoaminoglycans, and
 CC modulating wound healing. The peptide may also be used for blocking
 CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
 CC increase heparin half-life in circulation.
 XX
 XX Sequence 32 AA:
 SQ
 Query Match 100.0%; Score 36; DB 21; Length 32;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAAGA 8
 Db 1 ARKKAAGA 8
 |||||
 |||||

RESULT 9
 AAB08170
 ID AAB08170 standard; peptide; 40 AA.
 XX
 XX AAB08170;
 XX
 XX 04-DEC-2000 (first entry)
 XX
 XX Peptide modulating activity of heparin, and other glycans.
 DE
 XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
 KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
 KW cartilage differentiation; wound healing.
 XX
 XX Synthetic.
 OS
 XX WO200045831-A1.
 XX
 XX 10-AUG-2000.
 XX

PF 02-FEB-2000; 2000WO-US02853.
 XX
 XX 02-FEB-1999; 99US-0118276.
 XX
 XX (UYJE-) UNIV JEFFERSON THOMAS.
 XX
 XX San Antonio JD, Verrecchio A, Schick BP;
 XX
 XX WPI: 2000-543446/49.
 XX
 XX Novel synthetic peptides with high affinity for glycoaminoglycans and
 PT proteoglycans, useful for modulating heparin, promoting cell
 PT attachment, modulating tumour metastasis and modulating wound healing -
 PS Disclosure: Page 30; 76pp; English.
 XX
 XX The present sequence represents a synthetic peptide which has a high
 CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
 CC in methods for modulating heparin or other glycoaminoglycans with
 CC anticoagulant activity, promoting cell attachment or adhesion to
 CC natural or synthetic surfaces (especially vein grafts), modulating
 CC tumour cell metastasis, modulating cartilage differentiation, targeting
 CC drugs to epithelial cell surfaces (or to other cells expressing
 CC proteoglycans), modulating enzymes that act on glycoaminoglycan
 CC substrates, affinity purification of bioactive sequences of a
 CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
 CC anti-coagulant functions mediated through glycoaminoglycans, and
 CC modulating wound healing. The peptide may also be used for blocking
 CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
 CC increase heparin half-life in circulation.
 XX
 XX Sequence 40 AA:
 SQ
 Query Match 100.0%; Score 36; DB 21; Length 40;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAAGA 8
 Db 1 ARKKAAGA 8
 |||||
 |||||

RESULT 10
 AAB08144
 ID AAB08144 standard; peptide; 8 AA.
 XX
 XX AAB08144;
 XX
 XX 04-DEC-2000 (first entry)
 XX
 XX Peptide modulating activity of heparin, and other glycans.
 DE
 XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
 KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
 KW cartilage differentiation; wound healing.
 XX
 XX Synthetic.
 OS
 XX WO200045831-A1.
 XX
 XX 10-AUG-2000.
 XX
 XX 02-FEB-2000; 2000WO-US02853.
 XX
 XX 02-FEB-1999; 99US-0118276.
 XX
 XX (UYJE-) UNIV JEFFERSON THOMAS.
 XX

Key Location/Qualifiers
 FT Misc-difference 1.8
 FT /note= "this peptide may be repeated an
 FT unspecified number of times"

PI San Antonio JD, Verrecchio A, Schick BP;
XX
XX MPI: 2000-543446/49.
XX
XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
XX Disclosure: Page 23; 76pp: English.
XX
XX The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
XX
SQ Sequence 8 AA:

Query Match 91.7%; Score 33; DB 21; Length 8;
Best Local Similarity 87.5%; Pred. No. 7.8e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKKAKA 8
11:11111
Db 1 ARKKAKA 8

RESULT 11
AAB08165
ID AAB08165 standard; peptide: 23 AA.
XX
XX AAB08165;
AC
XX
XX 04-DEC-2000 (first entry)
DT
XX
XX Peptide modulating activity of heparin, and other glycans.
DE
XX
XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KM cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KM cartilage differentiation; wound healing.
KW
XX
XX Synthetic.
OS
XX
XX WO200045831-A1.
PN
XX
XX 10-AUG-2000.
PD
XX
XX 02-FEB-2000; 2000WO-US02853.
PE
XX
XX 02-FEB-1999; 99US-0118276.
PR
XX
XX (UYUE-) UNIV JEFFERSON THOMAS.
PS
XX
XX San Antonio JD, Verrecchio A, Schick BP;
PI
XX
XX MPI: 2000-543446/49.
DR
XX
XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
XX Disclosure: Page 24; 76pp: English.
PS
XX
XX The present sequence represents a synthetic peptide which has a high

CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
XX
SQ Sequence 23 AA:

Query Match 91.7%; Score 33; DB 21; Length 23;
Best Local Similarity 87.5%; Pred. No. 8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKKAKA 8
1111111
Db 1 ARKKAKA 8

RESULT 12
AAB08166
ID AAB08166 standard; peptide: 31 AA.
XX
XX AAB08166;
AC
XX
XX 04-DEC-2000 (first entry)
DT
XX
XX Peptide modulating activity of heparin, and other glycans.
DE
XX
XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KM cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KM cartilage differentiation; wound healing.
KW
XX
XX Synthetic.
OS
XX
XX WO200045831-A1.
PN
XX
XX 10-AUG-2000.
PD
XX
XX 02-FEB-2000; 2000WO-US02853.
PE
XX
XX 02-FEB-1999; 99US-0118276.
PR
XX
XX (UYUE-) UNIV JEFFERSON THOMAS.
PS
XX
XX San Antonio JD, Verrecchio A, Schick BP;
PI
XX
XX MPI: 2000-543446/49.
DR
XX
XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
XX Disclosure: Page 24; 76pp: English.
PS
XX
XX The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking

CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.

XX Sequence 31 AA;

Query Match 91.7%; Score 33; DB 21; Length 31;
Best Local Similarity 87.5%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAARA 8
Db 1 ARKKAARA 8

RESULT 13
AAW02564
ID AAW02564 standard; peptide: 14 AA.

AC AAW02564;
DT 13-JAN-1997 (first entry)
DE AKAP79 (residues 37-50).

XX Autoimmune disease; PACT59; PACT74; PACT36; PACT60; murine; T-cell; PKA;
KW mouse; A-kinase anchoring protein 79; CAMP-dependent protein kinase;
KW postsynaptic density; AKAP79; human forebrain; transcriptional activator;
KW interleukin 2; T-cell activation; calcineurin; T-cell clonal expansion;
KW calcium/calmodulin dependent protein phosphatase; T-cell response;
KW autoimmune related disease; therapy; immune response.

XX Synthetic.

OS
FH Key Location/Qualifiers
FT Misc-difference 2 /note="any amino acid"
XX

PN W09616172-A2.

PD 30-MAY-1996.

PF 22-NOV-1995; 95WO-US16039.

XX 17-JUL-1995; 95US-0503226.

PR 23-NOV-1994; 94US-0344227.

PR 15-MAR-1995; 95US-0404731.

XX (ICOS-) ICOS CORP.
PA (OREG-) STATE OF OREGON.
PA (UYOR-) UNIV OREGON HEALTH SCI.

PI Coghlan WM, Gallatin WM, Howard ML, Lockerbie RO;
PI Scott JD;

XX WPI; 1996-266608/27.

PT New modulators of anchoring protein function - used to develop
PT prods. for use in the treatment of auto-immune-related conditions.

XX Example 11; Page 34; 74pp; English.

XX This sequence represents residues 37-50 of the A-kinase anchoring
CC protein 79 (AKAP79). AKAP79 is responsible for anchoring CAMP-dependent
CC protein kinase (PKA) to specific intracellular sites. AKAP79 is
CC predominantly present in postsynaptic densities in the human forebrain.
CC The pathways that involve AKAP79 are important in many cell types and
CC have been implicated in many cell functions, including the
CC transcriptional activation of the interleukin 2 gene that is important in
CC T-cell activation. AKAP also binds to calcineurin (see AAW02536), which
CC is a calcium/calmodulin dependent protein phosphatase associated with
CC T-cell activation. By binding both PKA and calcineurin, AKAP79
CC co-localises a kinase and a phosphatase which may regulate flux through a
CC specific signalling pathway. The AKAP79 binding sequences can be used to

CC develop products for use in the treatment of autoimmune related
CC conditions. The AKAP79 binding proteins can be used in methods for
CC stimulating an immune response, and for stimulating activated T-cells for
CC selected clonal expansion. The proteins can also be used in a method for
CC enhancing T-cell responses to experimental stimuli for evaluation of
CC early events in T-cell biology and activation of the immune response.

XX Sequence 14 AA;

Query Match 88.9%; Score 32; DB 17; Length 14;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RKKAKA 8
Db 4 RKKAKA 10

RESULT 14
AAB14904
ID AAB14904 standard; Peptide: 14 AA.

AC AAB14904;

DT 08-JAN-2001 (first entry)

DE Human AKAP 79 peptide.

XX Human; calcineurin-binding peptide; A-kinase anchor protein; AKAP;
KW AKAP 79; immunostimulant; interleukin 2 expression modulation;
KW graft rejection; transplantation; T cell-mediated disorder.

XX Homo sapiens.

OS US6107104-A.

PN 22-AUG-2000.

PD 27-SEP-1996; 96US-0721458.

PF 23-NOV-1994; 94US-0344227.

PR 15-MAR-1995; 95US-0404731.

PR 17-JUL-1995; 95US-0503226.

XX (ICOS-) ICOS CORP.

PI Lockerbie RO, Gallatin WM, Lai Y, Howard ML;

PI WPI; 2000-578541/54.
PT Novel calcineurin deletion mutant having calcineurin polypeptide
PT sequence and binding A-kinase anchor proteins, for treating graft
PT rejection following organ transplantation and T cell-mediated disorders

XX Example 11; Column 21; 53pp; English.

XX The present sequence is a peptide derived from A-kinase anchor protein 79
CC (AKAP 79). AKAP 79 binds both CAMP-dependent protein kinase (PKA) and
CC calcineurin and so co-localises a kinase and a phosphatase that
CC may regulate flux through a specific signalling pathway. Calcineurin is a
CC Ca2+/calmodulin-dependent protein phosphatase which is involved in many
CC intracellular signalling pathways. It participates in regulation of IL-2
CC expression following T cell stimulation in T cells. Calcineurin-binding
CC peptides derived from AKAP 79 may be used to inhibit calcineurin activity
CC in a cell. The peptides are useful for treating graft rejection following
CC organ transplantation and for treating T cell-mediated disorders.
CC Calcineurin deletion mutants which bind AKAP 79 are useful for defining
CC an AKAP 79 binding site, for stimulating the immune response, stimulating
CC activated T cells for selected clonal expansion, or for enhancing T cell
CC responses to experimental stimuli for evaluation of early events in
CC T cell biology and activation of the immune response.

SQ Sequence 14 AA:
Query Match 88.9%; Score 32; DB 21; Length 14;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RKKAKA 8
Db 4 RKKAKA 10

RESULT 15
AAG17771
ID AAG17771 standard; Protein; 331 AA.
XX
AC AAG17771;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 18918.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 09-MAR-1999; 99US-0123180.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139494.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:19:38 ; Search time 10.4828 Seconds
(without alignments)
22.454 Million cell updates/sec

Title: US-09-496-391-5

Perfect score: 36

Sequence: 1 ARKKAAGA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	14	2	US-08-503-226B-38 Sequence 38, Appl
2	36	100.0	14	3	US-08-721-458B-38 Sequence 38, Appl
3	32	88.9	14	2	US-08-503-226B-37 Sequence 37, Appl
4	32	88.9	14	3	US-08-721-458B-37 Sequence 37, Appl
5	29	80.6	183	4	US-09-074-658B-77 Sequence 77, Appl
6	28	77.8	22	1	US-08-769-309A-23 Sequence 23, Appl
7	28	77.8	22	3	US-08-994-570-23 Sequence 23, Appl
8	28	77.8	180	6	5273901-7 Patent No. 5273901
9	28	77.8	180	6	5482709-6 Patent No. 5482709
10	28	77.8	299	2	US-08-923-856-1 Sequence 1, Appl
11	28	77.8	299	3	US-09-216-294-1 Sequence 10, Appl
12	27	75.0	17	2	US-09-115-209-10 Sequence 1, Appl
13	27	75.0	26	1	US-08-231-730A-46 Sequence 46, Appl
14	27	75.0	26	2	US-08-505-486-51 Sequence 46, Appl
15	27	75.0	26	3	US-08-689-489C-46 Sequence 46, Appl
16	27	75.0	26	3	US-08-801-028-51 Sequence 51, Appl
17	27	75.0	26	3	US-09-340-154-51 Sequence 51, Appl
18	27	75.0	26	4	US-09-232-802A-46 Sequence 46, Appl
19	27	75.0	26	4	US-09-482-611B-51 Sequence 51, Appl
20	27	75.0	26	5	PCT-US95-04718-46 Sequence 46, Appl
21	27	75.0	26	5	PCT-US95-09338-51 Sequence 51, Appl
22	27	75.0	26	5	PCT-US95-09339-51 Sequence 51, Appl
23	27	75.0	31	4	US-08-995-172-2 Sequence 2, Appl
24	27	75.0	31	4	US-08-839-674-28 Sequence 28, Appl
25	27	75.0	31	4	US-09-150-812-28 Sequence 28, Appl
26	27	75.0	31	4	US-09-041-889-41 Sequence 16, Appl
27	27	75.0	60	1	US-08-346-849-16 Sequence 16, Appl

28	27	75.0	60	2	US-08-293-284A-16 Sequence 16, Appl
29	27	75.0	61	1	US-08-346-849-17 Sequence 17, Appl
30	27	75.0	61	2	US-08-293-284A-17 Sequence 17, Appl
31	27	75.0	158	3	US-09-041-889-40 Sequence 40, Appl
32	27	75.0	226	3	US-09-041-889-32 Sequence 32, Appl
33	27	75.0	345	1	US-08-745-269-2 Sequence 2, Appl
34	27	75.0	365	2	US-08-428-243-9 Sequence 9, Appl
35	27	75.0	365	5	PCT-US93-10301-9 Sequence 9, Appl
36	27	75.0	404	2	US-08-428-243-7 Sequence 7, Appl
37	27	75.0	404	5	PCT-US93-10301-7 Sequence 7, Appl
38	27	75.0	422	1	US-07-996-772A-12 Sequence 12, Appl
39	27	75.0	435	2	US-08-031-538-11 Sequence 11, Appl
40	27	75.0	445	2	US-08-157-185-2 Sequence 2, Appl
41	27	75.0	445	3	US-08-281-526B-2 Sequence 2, Appl
42	27	75.0	445	4	US-09-450-797-2 Sequence 2, Appl
43	27	75.0	445	4	US-09-328-314-16 Sequence 16, Appl
44	27	75.0	445	4	US-09-450-790A-2 Sequence 2, Appl
45	27	75.0	445	4	US-09-332-837-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-503-226B-38
Sequence 38, Application US/08503226B
Patent No. 5871945
GENERAL INFORMATION:
APPLICANT: Lockerbie, Robert Owen, et al.
TITLE OF INVENTION: Calcineurin Inhibitory Compounds and
ANCHORING PROTEIN
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/503,226B
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/404,731
FILING DATE: 15-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,227
FILING DATE: 23-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32861
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-503-226B-38
Query Match 100.0%; Score 36; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKAKA 8
Db 3 ARKAKA 10

RESULT 2

US-08-721-458B-38
; Sequence 38, Application US/08721458B
; Patent No. 6107104
; GENERAL INFORMATION:
; APPLICANT: Lockerbile, Robert Owen, et al.
; TITLE OF INVENTION: Calcineurin Inhibitory Compounds
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,458B
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/404,731
; FILING DATE: 15-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,227
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/503,226
; FILING DATE: 17-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33276
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-4740448
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-721-458B-38

Query Match 100.0%; Score 36; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKAKA 8
Db 3 ARKAKA 10

RESULT 3

US-08-503-226B-37
; Sequence 37, Application US/08503226B
; Patent No. 5871945
; GENERAL INFORMATION:
; APPLICANT: Lockerbile, Robert Owen, et al.
; TITLE OF INVENTION: Calcineurin Inhibitory Compounds and
; ANCHORING PROTEIN

NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/503,226B
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/404,731
; FILING DATE: 15-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,227
; FILING DATE: 23-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32861
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-4740448
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-503-226B-37

Query Match 88.9%; Score 32; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RKKAKA 8
Db 4 RKKAKA 10

RESULT 4

US-08-721-458B-37
; Sequence 37, Application US/08721458B
; Patent No. 6107104
; GENERAL INFORMATION:
; APPLICANT: Lockerbile, Robert Owen, et al.
; TITLE OF INVENTION: Calcineurin Inhibitory Compounds
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,458B
; FILING DATE:
; CLASSIFICATION: 436

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/404,731
FILING DATE: 15-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,227
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/503,226
FILING DATE: 17-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/33276
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-6300
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-721-458b-37

Query Match 88.9%; Score 32; DB 3; Length 14;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RKKAKA 8
111111
Db 4 RKKAKA 10

RESULT 5
US-09-074-658-77
Sequence 77, Application US/09074658
Patent No. 6184371
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Run-Pan Du
APPLICANT: Quljun Wang
APPLICANT: Yang, Yan-ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,658
FILING DATE: 08-MAY-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 183 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-658-77

Query Match 80.6%; Score 29; DB 4; Length 183;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKAKA 8
111111
Db 37 ARKAKA 44

RESULT 6
US-08-769-309A-23
Sequence 23, Application US/08769309A
Patent No. 5741890
GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Nauert, Brian J.,
APPLICANT: Klauck, Theresa M.
TITLE OF INVENTION: Protein Binding Domains of Gravin
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/769,309A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5741890and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-769-309A-23

Query Match 77.8%; Score 28; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RKKAK 7
111111
Db 10 RKKAK 15

RESULT 7
US-08-994-570-23
Sequence 23, Application US/08994570
Patent No. 6090929
GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Nauert, Brian J.,
APPLICANT: Klauck, Theresa M.
TITLE OF INVENTION: Protein Binding Domains of Gravin

NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/994,570
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: NO. 6090929and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-994-570-23

Query Match 77.8%; Score 28; DB 3; Length 22;
Best local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RRRRAK 7
Db 10 RRRRAK 15

RESULT 8
5273901-7
Patent No. 5273901
APPLICANT: JACOBSON, JAMES W.; STRAUSSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
SPOROZOITE 21.5 KB ANTIGEN, AC-6B
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/581,693
FILING DATE: 12-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1988
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 7;
LENGTH: 180
5273901-7

Query Match 77.8%; Score 28; DB 6; Length 180;
Best local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKAKAKA 8
Db 161 AOEKAKAKA 168

RESULT 9
5482709-6
Patent No. 5482709
APPLICANT: JACOBSON, JAMES W.; STRAUSSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
NUMBER OF SEQUENCES: 10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/148,432
FILING DATE: 08-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 581,693
FILING DATE: 12-SEP-1990
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1989
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 6;
LENGTH: 180
5482709-6

Query Match 77.8%; Score 28; DB 6; Length 180;
Best local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKAKAKA 8
Db 161 AOEKAKAKA 168

RESULT 10
US-08-923-856-1
Sequence 1, Application US/08923856
Patent No. 5928894
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Tang, Tom
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN ACTVA-ORF4-LIKE PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,856
FILING DATE: Filed Herewith
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0380 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 299 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PANCNOT01
CLONE: 223909
US-08-923-856-1

Query Match 77.8%; Score 28; DB 2; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RKKA 7
Db 40 RKKA 45

RESULT 11
US-09-216-294-1
Sequence 1, Application US/09216294
Patent No. 6080723
GENERAL INFORMATION:
APPLICANT: Lai, Preeti
APPLICANT: Tang, Tom
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN ACTVA-ORF4-LIKE PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/216,294
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/923,856
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0380 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 299 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PANCNOT01
CLONE: 223909
US-09-216-294-1

Query Match 77.8%; Score 28; DB 3; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RKKA 7
Db 40 RKKA 45

RESULT 12
US-09-115-209-10
Sequence 10, Application US/09115209
Patent No. 5998375
GENERAL INFORMATION:
APPLICANT: Thogersen, Henning
APPLICANT: Madsen, Kjeld
APPLICANT: Olsen, Uffe B.
APPLICANT: Johansen, Nils L.
APPLICANT: Scheidegger, Mark
TITLE OF INVENTION: No. 59983751ceptin Analogues
FILE REFERENCE: 5285.200-US
CURRENT APPLICATION NUMBER: US/09/115,209
CURRENT FILING DATE: 1998-07-14
EARLIER APPLICATION NUMBER: 0867/97
EARLIER FILING DATE: 1997-07-15
EARLIER APPLICATION NUMBER: 60/052,862
EARLIER FILING DATE: 1997-07-17
NUMBER OF SEQ ID NOS: 76
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 10
LENGTH: 17
TYPE: PRT
ORGANISM: Human
US-09-115-209-10

Query Match 75.0%; Score 27; DB 2; Length 17;
Best Local Similarity 85.7%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKKA 7
Db 7 ARKKA 13

RESULT 13
US-08-231-730A-46
Sequence 46, Application US/08231730A
Patent No. 5561107
GENERAL INFORMATION:
APPLICANT: JAYNES, JESSE M.
APPLICANT: JULIAN, GORDON R.
TITLE OF INVENTION: METHOD OF ENHANCING WOUND HEALING BY STIMULATING FIBROBLAST
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: STEVEN J. HULTQUIST
ADDRESS: INTELLECTUAL PROPERTY/TECHNOLOGY LAW
STREET: 200 PARK DRIVE, SUITE 210
STREET: P.O. BOX 14329
CITY: RESEARCH TRIANGLE PARK
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
COMPUTER: APPLE MACINTOSH
OPERATING SYSTEM: MACINTOSH
SOFTWARE: M.S. WORD 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,730A
FILING DATE: 04-20-94
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,476
FILING DATE: 04-08-94
APPLICATION NUMBER: 08/039,620
FILING DATE: 06-04-93
APPLICATION NUMBER: 08/148,491
FILING DATE: 11-08-93
APPLICATION NUMBER: 08/148,889
FILING DATE: 11-08-93
ATTORNEY/AGENT INFORMATION:
NAME: HULTQUIST, STEVEN J.

REGISTRATION NUMBER: 28021
REFERENCE/DOCKET NUMBER: 4013-106
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)990-9531
TELEFAX: (919)990-9532
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 26
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
HYPOTHETICAL: NO
COMPLETE PEPTIDE
ORIGINAL SOURCE: SYNTHETIC
IMMEDIATE SOURCE: SYNTHETIC
PUBLICATION INFORMATION: NOT PREVIOUSLY PUBLISHED
US-08-231-730A-46

Query Match 75.0%; Score 27; DB 1; Length 26;
Best Local Similarity 85.7%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKAKK 7
Db 7 ARKAKK 13

RESULT 14
US-08-505-486-51
Sequence 51, Application US/08505486
Patent No. 5955573
GENERAL INFORMATION:
APPLICANT: Jesse M. Jaynes
TITLE OF INVENTION: UBIQUITIN-LYTIC PEPTIDE FUSION GENE
TITLE OF INVENTION: CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
TITLE OF INVENTION: METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
STREET: 555 Thirteenth Street N.W.
CITY: Washington
STATE: D. C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
COMPUTER: IBM COMPATIBLE
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/505,486
FILING DATE: 21-JUL-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/279,472
FILING DATE: 22-JUL-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: WALKER, BARBARA W.
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 2093-117A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 26
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PEPTIDE
DESCRIPTION: NO
HYPOTHETICAL: NO

FRAGMENT TYPE: COMPLETE PEPTIDE
ORIGINAL SOURCE: SYNTHETIC
IMMEDIATE SOURCE: SYNTHETIC
PUBLICATION INFORMATION: NOT PREVIOUSLY PUBLISHED
US-08-505-486-51

Query Match 75.0%; Score 27; DB 2; Length 26;
Best Local Similarity 85.7%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKAKK 7
Db 7 ARKAKK 13

RESULT 15
US-08-689-489C-46
Sequence 46, Application US/08689489C
Patent No. 6001805

GENERAL INFORMATION:
APPLICANT: Jesse M. Jaynes, Gordon R. Julian
TITLE OF INVENTION: Method of Enhancing Wound Healing By
Stimulating Fibroblast and Keratinocyte Growth In
TITLE OF INVENTION: Vivo, utilizing Amphipathic Reptides
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz
STREET: 555 13TH STREET
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/689,489C
FILING DATE: August 12, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/231,730
FILING DATE: April 20, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/225,476
FILING DATE: April 8, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,620
FILING DATE: June 4, 1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/148,889
FILING DATE: No. 6001805ember 8, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/148,491
FILING DATE: No. 6001805ember, 8, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Mark I. Bowditch
REGISTRATION NUMBER: 40,315
REFERENCE/DOCKET NUMBER: 2093-120
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-783-6040
TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
US-08-689-489C-46

Query Match 75.0%; Score 27; DB 3; Length 26;

Best Local Similarity 85.78; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Oy 1 ARKRAK 7
| | | | |
Db 7 ARKRAK 13

Search completed: April 8, 2003, 11:31:14
Job time : 12.4828 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:28:39 ; Search time 11.0345 Seconds
(without alignments)
44.324 Million cell updates/sec

Title: US-09-496-391-5
Perfect score: 36
Sequence: 1 ARKKAKA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB pep: *
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB pep: *
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB pep: *
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB pep: *
5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB pep: *
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB pep: *
7: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB pep: *
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB pep: *
9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB pep: *
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB pep: *
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB pep: *
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB pep: *
13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB pep: *
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	83.3	14	10 US-09-765-086-202	Sequence 202, App
2	29	80.6	43	10 US-09-864-761-45922	Sequence 45922, A
3	29	80.6	448	10 US-09-815-242-12017	Sequence 12017, A
4	28	77.8	157	10 US-09-764-853-525	Sequence 525, App
5	27	75.0	110	9 US-09-925-289-1242	Sequence 1242, App
6	27	75.0	110	10 US-09-925-299-1242	Sequence 1242, App
7	27	75.0	228	10 US-09-815-242-11542	Sequence 11542, A
8	27	75.0	269	9 US-10-027-806-6	Sequence 6, Appli
9	27	75.0	269	9 US-10-034-623-6	Sequence 6, Appli
10	27	75.0	269	9 US-10-027-801-6	Sequence 6, Appli
11	27	75.0	272	9 US-09-738-626-3538	Sequence 3538, App
12	27	75.0	304	9 US-09-738-626-5018	Sequence 5018, App
13	27	75.0	371	9 US-09-738-626-5471	Sequence 5471, App
14	27	75.0	412	9 US-09-981-353-87	Sequence 87, Appli
15	27	75.0	421	9 US-10-012-055-2	Sequence 2, Appli
16	27	75.0	442	10 US-09-925-300-1594	Sequence 1594, App
17	27	75.0	445	10 US-09-989-861-16	Sequence 16, Appli
18	27	75.0	647	9 US-09-991-262-50	Sequence 50, Appli
19	27	75.0	675	9 US-09-991-262-52	Sequence 52, Appli

20	26	72.2	93	9 US-10-091-504-1103	Sequence 1103, App
21	26	72.2	93	10 US-09-764-869-1103	Sequence 1103, App
22	26	72.2	113	10 US-09-815-242-5086	Sequence 5086, App
23	26	72.2	159	10 US-09-925-301-972	Sequence 972, App
24	26	72.2	354	10 US-09-886-055-331	Sequence 331, App
25	26	72.2	369	9 US-09-738-626-4885	Sequence 4885, App
26	26	72.2	406	10 US-09-938-540-2	Sequence 2, Appli
27	26	72.2	406	9 US-09-983-802-187	Sequence 187, App
28	26	72.2	502	9 US-10-027-806-34	Sequence 34, Appli
29	26	72.2	502	9 US-10-034-623-34	Sequence 34, Appli
30	26	72.2	502	9 US-10-027-801-34	Sequence 34, Appli
31	26	72.2	503	9 US-10-027-806-66	Sequence 66, Appli
32	26	72.2	503	9 US-10-034-623-66	Sequence 66, Appli
33	26	72.2	503	9 US-10-027-801-66	Sequence 66, Appli
34	26	72.2	536	9 US-10-028-072-490	Sequence 490, App
35	26	72.2	536	9 US-10-121-049-490	Sequence 490, App
36	26	72.2	536	9 US-10-123-904-490	Sequence 490, App
37	26	72.2	536	9 US-10-140-470-490	Sequence 490, App
38	26	72.2	536	9 US-10-175-746-490	Sequence 490, App
39	26	72.2	536	9 US-10-176-918-490	Sequence 490, App
40	26	72.2	536	9 US-10-176-921-490	Sequence 490, App
41	26	72.2	536	9 US-10-137-865-490	Sequence 490, App
42	26	72.2	536	9 US-10-140-474-490	Sequence 490, App
43	26	72.2	536	9 US-10-142-431-490	Sequence 490, App
44	26	72.2	536	9 US-10-143-114-490	Sequence 490, App
45	26	72.2	536	9 US-10-140-002-490	Sequence 490, App

ALIGNMENTS

RESULT 1
US-09-765-086-202
Sequence 202, Application US/09765086
Patent No. US20010046498A1
GENERAL INFORMATION:
APPLICANT: Ruoslantti, Erkki
APPLICANT: Pasqualini, Renata
APPLICANT: Madh, Arad
APPLICANT: Bredesen, Dale E.
APPLICANT: Ellerby, H. Michael
TITLE OF INVENTION: Chimeric Prostate-Homing Peptides With
FILE REFERENCE: P-LJ 3844
CURRENT APPLICATION NUMBER: US/09/765,086
CURRENT FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US 09/489,582
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 235
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 202
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: synthetic peptide
US-09-765-086-202
Query Match 83.3%; Score 30; DB 10; Length 14;
Best Local Similarity 87.5%; Pred. No. 4;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Cy 1 ARKKAKA 8
Db 2 ARKKAKA 9
RESULT 2
US-09-864-761-45922
Sequence 45922, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.

```
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Weisheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 45922
LENGTH: 43
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC011155.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.44
OTHER INFORMATION: EST_HUMAN HIT: AW85305.1, EVALU0 9.20e+00
US-09-864-761-45922

Query Match
Best Local Similarity 80.6%; Score 29; DB 10; Length 43;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 RKKAKA 8
    |||||
DB 3 RKKASKA 9
```

```
RESULT 3
US-09-815-242-12017
; Sequence 12017, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
```

```
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELIPRA.011a
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12017
LENGTH: 448
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12017

Query Match
Best Local Similarity 80.6%; Score 29; DB 10; Length 448;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 RKKAKA 8
    |||||
DB 347 RKKAKA 353
```

```
RESULT 4
US-09-764-853-525
; Sequence 525, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 525
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (88)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-853-525
```

```
Query Match
Best Local Similarity 77.8%; Score 28; DB 10; Length 157;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 RKKAAK 7
    |||||
DB 77 RKKAAK 82
```

RESULT 5
US-09-925-299-1242
; Sequence 1242, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1242
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (103)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1242

Query Match 75.0%; Score 27; DB 9; Length 110;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ARKAKA 8
||||| 11
DB 18 ARKKRTKA 25

RESULT 6
US-09-925-299-1242
; Sequence 1242, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1242
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (103)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1242

Query Match 75.0%; Score 27; DB 10; Length 110;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ARKAKA 8
||||| 11
DB 18 ARKKRTKA 25

RESULT 7

US-09-815-242-11542
; Sequence 11542, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11542
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11542

Query Match 75.0%; Score 27; DB 10; Length 228;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 KKAKA 8
||||| 11
DB 12 KKAKA 17

RESULT 8
US-10-027-806-6
; Sequence 6, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCCP-002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-6

Query Match 75.0%; Score 27; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKKA 6
DB 128 ARKKA 133

RESULT 9
US-10-034-623-6

; Sequence 6, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-6

Query Match 75.0%; Score 27; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKKA 6
DB 128 ARKKA 133

RESULT 10
US-10-027-801-6

; Sequence 6, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-6

Query Match 75.0%; Score 27; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKKA 6
DB 128 ARKKA 133

RESULT 11
US-09-738-626-3538
; Sequence 3538, Application US/09738626

; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3538
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3538

Query Match 75.0%; Score 27; DB 9; Length 272;
Best Local Similarity 75.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKKA 8
DB 111 ARKKA 118

RESULT 12
US-09-738-626-5018

; Sequence 5018, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5018
; LENGTH: 304
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5018

Query Match 75.0%; Score 27; DB 9; Length 304;

Best Local Similarity 75.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ARKRAAK 8
||| |
Db 244 ARKSAVKA 251

RESULT 13

US-09-738-626-5471
; Sequence 5471, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAKOTO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent ver. 3.0
; SEQ ID NO 5471
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5471

Query Match 75.0%; Score 27; DB 9; Length 371;
Best Local Similarity 85.7%; Pred. No. 4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKRAAK 7
| | | | |
Db 287 ARKRAAK 293

RESULT 14

US-09-981-353-87
; Sequence 87, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 87
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 1543330CD1
US-09-981-353-87

Query Match 75.0%; Score 27; DB 9; Length 412;

Best Local Similarity 85.7%; Pred. No. 4.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKRAAK 7
||| |
Db 47 ARKRAAK 53

RESULT 15

US-10-012-055-2
; Sequence 2, Application US/10012055
; Patent No. US20020164750A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: 56634, A NOVEL HUMAN PHOSPHATIDYLINOSITOL 4-PHOSPHATE 5-KINASE
; FILE REFERENCE: 10448-114001
; CURRENT APPLICATION NUMBER: US/10/012,055
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 60/248,325
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-055-2

Query Match 75.0%; Score 27; DB 9; Length 421;
Best Local Similarity 75.0%; Pred. No. 4.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKRAAK 8
| : | | | |
Db 381 ARKRAAK 388

Search completed: April 8, 2003, 11:52:43
Job time : 13.0345 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:18:53 ; Search time 12.6897 Seconds
(without alignments)
60.606 Million cell updates/sec

Title: US-09-496-391-5
Perfect score: 36
Sequence: 1 ARKKAAGA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	91.7	187	T50946	hypothetical prote
2	33	91.7	473	T32326	hypothetical prote
3	32	88.9	333	T37871	hypothetical nucle
4	32	88.9	427	A43453	A-kinase anchor pr
5	32	88.9	428	A42311	hypothetical prote
6	32	88.9	478	T09896	hypothetical prote
7	32	88.9	1741	T15978	hypothetical prote
8	31	86.1	104	B39435	hypothetical prote
9	31	86.1	105	B69333	conserved hypotet
10	30	83.3	166	A13415	hypothetical prote
11	30	83.3	231	S59589	histone H1 - Chlam
12	30	83.3	267	S76499	hypothetical prote
13	30	83.3	781	JC7382	DNA-directed DNA p
14	30	83.3	924	T06636	hypothetical prote
15	29	80.6	97	B97662	hypothetical prote
16	29	80.6	139	I37781	Ig variable region
17	29	80.6	144	S69345	oviduct-specific p
18	29	80.6	212	A28470	histone H1 - mouse
19	29	80.6	227	B81054	deda protein, prob
20	29	80.6	277	T34625	probable NLP/p60 f
21	29	80.6	309	T21195	hypothetical prote
22	29	80.6	324	D71417	cytochrome P450 d1
23	29	80.6	352	T36719	probable integral
24	29	80.6	378	T25558	hypothetical prote
25	29	80.6	448	A83091	tryptophanyl-trna
26	29	80.6	488	T05313	hypothetical prote
27	29	80.6	634	T08810	probable ABC-type
28	29	80.6	640	B96784	hypothetical prote
29	29	80.6	647	T39141	hypothetical prote

30	29	80.6	741	2	D81798
31	29	80.6	783	2	A48998
32	29	80.6	855	2	A48168
33	29	80.6	900	2	S70630
34	29	80.6	1293	2	T01512
35	28	77.8	66	2	A42814
36	28	77.8	66	2	C97593
37	28	77.8	78	2	T32171
38	28	77.8	141	2	T45687
39	28	77.8	165	2	S75414
40	28	77.8	224	2	S16259
41	28	77.8	239	2	T34945
42	28	77.8	254	2	F97841
43	28	77.8	298	2	D90162
44	28	77.8	319	2	A41773
45	28	77.8	338	2	A00278

ALIGNMENTS

```

RESULT 1
T50946
hypothetical protein B24P7.10 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 28-Jul-2000
C:Accession: T50946
R:Schulte, U.; Algin, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, July 2000
A:Reference number: 225286
A:Accession: T50946
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-187 <SCH>
A:Cross-references: EMBL:AL389890; GSPDB:GN00116; NCSP:B24P7.10
A:Experimental source: BMC clone B24P7; strain OR74A
C:Genetics:
A:Gene: NCSP:B24P7.10
A:Map position: 6
A:Introns: 14/1; 20/3; 122/3
C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPBC16C6.05

Query Match
Best Local Similarity 87.5%; Score 33; DB 2; Length 187;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKKAAGA 8
Db 71 ARKKAAGA 78

RESULT 2
T32326
hypothetical protein C41H7.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32326
R:Geisels, C.; Wamsley, P.; Elliott, G.; Smith, A.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid C41H7.
A:Reference number: 221150
A:Accession: T32326
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-473 <GEI>
A:Cross-references: EMBL:AF025450; PIDN:AAB70938.1; GSPDB:GN00020; CESP:C41H7.7
A:Experimental source: strain Bristol N2; clone C41H7
C:Genetics:
A:Gene: CESP:C41H7.7
A:Map position: 2
A:Introns: 64/3; 110/1; 150/1; 183/2

Query Match
Best Local Similarity 91.7%; Score 33; DB 2; Length 473;

```

Best Local Similarity 87.5%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKA 8
DB 199 SRKKA 206

RESULT 3

T37871
hypothetical nucleolar protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37871
R:Skellon, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A:Reference number: 221751
A:Accession: T37871
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-333 <SKE>
A:Cross-references: EMBL:298597; PIDN:CAB11214.1; GSPDB:GN00066; SPDB:SPAC17H9.05
A:Experimental source: strain 972h-; cosmid c17H9
C:Genetics:
A:Gene: SPDB:SPAC17H9.05
A:Map position: 1

Query Match 88.9%; Score 32; DB 2; Length 333;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKA 8
DB 11 AOKKA 18

RESULT 4

A43453
A-kinase anchor protein 79 - human
N:Alternate names: AKAP 79; cAMP-dependent protein kinase RII subunit-binding protein
C:Species: Homo sapiens (man)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C:Accession: A43453
R:Carry, D.W.; Stofo-Hahn, R.E.; Fraser, I.D.; Cone, R.D.; Scott, J.D.
J. Biol. Chem. 267, 16816-16823, 1992
A:Title: Localization of the cAMP-dependent protein kinase to the postsynaptic densities
A:Reference number: A43453; MID:92380978; PMID:1512224
A:Accession: A43453
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <CAR>
A:Cross-references: GB:M90359; NID:g178323; PID:g178324
A:Experimental source: thyroid
A:Note: sequence extracted from NCBI backbone (NCBIN:111869, NCBI:P.111870)

Query Match 88.9%; Score 32; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RKKAKA 8
DB 40 RKKAKA 46

RESULT 5

A42311
A-kinase anchor protein 75 - bovine
N:Alternate names: cAMP-dependent protein kinase II beta chain-binding protein
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A42311; A39782
R:Hirsch, A.H.; Glantz, S.B.; Li, Y.; You, Y.; Rubin, C.S.
J. Biol. Chem. 267, 2131-2134, 1992

A:Title: Cloning and expression of an intron-less gene for AKAP 75, an anchor protein
A:Reference number: A42311; MID:92129278; PMID:173921

A:Accession: A42311
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <HIR>
A:Cross-references: GB:M82914; NID:g162637; PIDN:AAA30366.1; PID:g162638
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIN:78206, NCBI:P.78208)
R:Bregman, D.B.; Hirsch, A.H.; Rubin, C.S.
J. Biol. Chem. 266, 7207-7213, 1991
A:Title: Molecular characterization of bovine brain P75, a high affinity binding prot
A:Reference number: A39782; MID:91201382; PMID:2016523
A:Accession: A39782
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 171-224, 'T', 226-428 <BRE>
A:Cross-references: GB:M60292; NID:g163474; PIDN:AAA30682.1; PID:g163475
C:Keywords: phosphoprotein

Query Match 88.9%; Score 32; DB 2; Length 428;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RKKAKA 8
DB 40 RKKAKA 46

RESULT 6

T09896
hypothetical protein T22A6.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 24-Nov-1999
R:Bevan, M.; Zimmermann, W.; Gruenisen, A.; Wambolt, R.; Bancroft, I.; Mewes, H.W.;
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16896
A:Accession: T09896
A:Molecule type: DNA
A:Residues: 1-478 <BEV>
A:Cross-references: EMBL:AL078637; GSPDB:GN00062; ATSP:T22A6.160
A:Experimental source: cultivar Columbia; BAC clone T22A6
C:Genetics:
A:Gene: ATSP:T22A6.160
A:Map position: 4
A:Introns: 334/3; 402/3
C:Superfamily: Arabidopsis thaliana hypothetical protein T22A6.160

Query Match 88.9%; Score 32; DB 2; Length 478;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKA 7
DB 278 ARKKA 284

RESULT 7

T15978
hypothetical protein F08F8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 23-Mar-2001
C:Accession: T15978
R:Du, Z.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F08F8.
A:Reference number: S61147
A:Accession: T15978
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1741 <DU>
A:Cross-references: EMBL:U28991; NID:g861364; PID:g861366; PIDN:AAA68385.1; CESP:F08F

A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F08P8.4
A:introns: 33/1; 64/1; 130/3; 201/2; 251/3; 362/2; 568/2; 628/3; 648/3; 655/3; 669/3; 73
/2; 1676/3

Query Match 88.9%; Score 32; DB 2; Length 1741;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKKAAR 7
| | | | |
Db 368 ARKKAAR 374

RESULT 8
B39435
hypothetical protein (mannitol phosphotransferase system region) - Enterococcus faecalis
C:Species: Enterococcus faecalis
C>Date: 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change 18-Jun-1993
C:Accession: B39435
R:Fischer, R.; von Strandmann, R.P.; Hengstenberg, W.
J. Bacteriol. 173, 3709-3715, 1991
A:Title: Mannitol-specific phosphoenolpyruvate-dependent phosphotransferase system of En
phosphate dehydrogenase gene, expression in Escherichia coli, and comparison of the gene
A:Reference number: A39435; MUID:91267934; PMID:1504856
A:Accession: B39435
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-104 <FIS>
A:Cross-references: GB:M38386

Query Match 86.1%; Score 31; DB 2; Length 104;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKKAAR 8
| | | | |
Db 88 ARKKAAR 95

RESULT 9
B69333
conserved hypothetical protein AF0666 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: B69333
R:Klen, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterlind, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: B69333
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-105 <KLE>
A:Cross-references: GB:AE001059; GB:AE000782; NID:92689382; PIDN:AAB90575.1; PID:9264995
C:Superfamily: hypothetical protein MJ1243

Query Match 86.1%; Score 31; DB 1; Length 105;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKKAAR 8
| | | | |
Db 80 ARKKAAR 87

RESULT 10
A13415

hypothetical protein BME11311 [Imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: A13415
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Muje, C.; Los, T.; Ivanov
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella meli
A:Reference number: AD3252; PMID:11756688
A:Accession: A13415
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-166 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL52492.1; PID:917983301; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11311
A:Map position: 1

Query Match 83.3%; Score 30; DB 2; Length 166;
Best Local Similarity 87.5%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKKAAR 8
| | | | |
Db 28 ARKKAAR 35

RESULT 11
S59589
histone H1 - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 23-Jul-1999
C:Accession: S59589; S62122
R:Febr, S.; Mueller, K.; Lindauer, A.; Park, P.B.; Cornelius, T.; Schmitt, R.
Curr. Genet. 28, 333-345, 1995
A:Title: The organization structure and regulatory elements of Chlamydomonas histone
A:Reference number: S59581; MUID:96120862; PMID:8590479
A:Accession: S59589
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-231 <FAB>
A:Cross-references: EMBL:U16726
A:Note: the authors did not translate the codon for residue 1
R:Febr, S.; Ford, C.; Lindauer, A.; Mueller, K.; Cornelius, T.; Schmitt, R.
submitted to the EMBL Data Library, October 1994
A:Description: The organization, structure and controlling elements of Chlamydomonas
A:Reference number: S62122
A:Accession: S62122
A:Molecule type: DNA
A:Residues: 1-173, 'P', 174-231 <FAB>
A:Cross-references: EMBL:U16726; NID:9571479; PIDN:AAA98452.1; PID:9571480
C:Genetics:
A:introns: 62/3; 101/3
C:Superfamily: histone H1
C:Keywords: chromosomal protein; DNA binding; nucleosome

Query Match 83.3%; Score 30; DB 2; Length 231;
Best Local Similarity 87.5%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKKAAR 8
| | | | |
Db 14 ARKKAAR 21

RESULT 12
S76499
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S76499

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-130, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76499

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-267 <KAN>

A:Cross-references: EMBL:D64002; GB:AB001339; NID:g1001612; PIDN:BAAI0345.1; PID:g100161
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: probable lipid transfer protein M30

Query Match

Best Local Similarity 83.3%; Score 30; DB 2; Length 267;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKKA 8

Db 141 ARKKA 148

RESULT 13

JC7382

DNA-directed DNA polymerase (EC 2.7.7.7) B3 - *Sulfurisphaera ohwakensis*

N:Alternate names: DNA polymerase alpha, beta, gamma; DNA polymerase I, II, III

C:Species: *Sulfurisphaera ohwakensis*

C>Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 31-Dec-2000

C:Accession: JC7382

R:Iwai, T.; Kurosawa, N.; Itoh, Y.H.; Kimura, N.; Horiuchi, T.

DNA Res. 7, 243-251, 2000

A:Title: Sequence analysis of three family B DNA polymerases from the thermoacidophilic

A:Reference number: JC7380

A:Accession: JC7382

A:Molecule type: DNA

A:Residues: 1-781 <TMA>

A:Cross-references: DDBJ:AB022376

C:Comment: This enzyme has both 3'-5' exonuclease and polymerase activities, and plays a

C:Genetics:

A:Gene: B3

C:Keywords: DNA replication; exonuclease; metal binding; nucleotidyltransferase

Query Match

Best Local Similarity 83.3%; Score 30; DB 2; Length 781;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKKA 8

Db 696 ARKKA 703

RESULT 14

T06636

hypothetical protein T20K18.130 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999

C:Accession: T06636

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew

submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15790

A:Accession: T06636

A:Molecule type: DNA

A:Residues: 1-924 <BEV>

A:Cross-references: EMBL:AL049640; GSPDB:GN00062; ATSP:T20K18.130

A:Experimental source: cultivar Columbia; BAC clone T20K18

C:Genetics:

A:Gene: ATSP:T20K18.130

A:Map position: 4

A:Introns: 209/2; 699/3; 753/3; 785/2; 807/2; 853/3; 912/3

Query Match

Best Local Similarity 83.3%; Score 30; DB 2; Length 924;

Matches 7; Conservative 87.5%; Pred. No. 2.6e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKKA 8

Db 607 ARKKA 614

RESULT 15

B97662

hypothetical protein AGR_C.4578 [imported] - *Agrobacterium tumefaciens* (strain C58, C

C:Species: *Agrobacterium tumefaciens*

C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002

C:Accession: B97662

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz,

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium*

A:Reference number: A97359; PMID:11743194

A:Accession: B97662

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-97 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK88251.1; PID:g15157709; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C.4578

A:Map position: circular chromosome

Query Match

Best Local Similarity 80.6%; Score 29; DB 2; Length 97;

Matches 6; Conservative 85.7%; Pred. No. 66;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 RKA 8

Db 91 RKA 97

Search completed: April 8, 2003, 11:29:54

Job time : 16.6897 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:12:02 ; Search time 6.2069 Seconds
(without alignments)
53.458 Million cell updates/sec

Title: US-09-496-391-5
Perfect score: 36
Sequence: 1 ARKAKAKA 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt.40:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	32	88.9	EBP2_SCHPO	013802 schizosacch
2	32	88.9	AKA5_HUMAN	P24588 homo sapien
3	32	88.9	AKA5_BOVIN	P24275 bos taurus
4	30	83.3	6171_SYNY3	Q55707 synecocyst
5	30	83.3	4041_ISCS_METTE	P57795 methanosarc
6	29	80.6	2111_HI2_MOUSE	P15864 mus musculu
7	29	80.6	3611_SUR6_HUMAN	O75683 homo sapien
8	29	80.6	6231_ABF2_HUMAN	Q9ug63 homo sapien
9	29	80.6	8551_NOL1_HUMAN	P46087 homo sapien
10	29	90.0	XPC_MOUSE	P51612 mus musculu
11	28	77.8	RL29_AGRF5	Q8ue26 agrobacteri
12	28	77.8	661_RL29_RHIME	Q92qg2 rhizobium m
13	28	77.8	1651_RS13_SUIISO	P95986 sulfolobus
14	28	77.8	1791_RK6_GUTH	O46908 guillardia
15	28	77.8	2101_SODM_ASPEU	Q92450 aspergillus
16	28	77.8	2241_LEA3_WHEAT	Q03968 triticum ae
17	28	77.8	2991_NUSG_STRYG	P27309 streptomyce
18	28	77.8	3791_SWMY_YEAST	P04803 saccharomyc
19	28	77.8	3801_FTS2_MYCPN	P75464 mycoplasma
20	28	77.8	3841_DNAJ_HALME	Q9uhb8 halobacteri
21	28	77.8	3961_ISCS_HALME	O54055 halobacteri
22	28	77.8	4141_CDVI_MOUSE	O35055 ruminococcu
23	28	77.8	1217_SVY_FUGRU	P49696 fugu rubrip
24	28	77.8	1232_TOPI2_TRYCR	P30190 trypanosoma
25	28	77.8	1264_SVY2_HUMAN	P26640 homo sapien
26	27	75.0	801_SSS2_SCYCA	P11020 scyllorhinu
27	27	75.0	1381_RLI9_RICCN	Q92j5b rickettsia
28	27	75.0	1601_RS16_MYCLE	O33014 mycobacteri
29	27	75.0	1621_RS16_MYCTU	O10795 mycobacteri
30	27	75.0	1711_H1_EGCHR	P02257 echinolampa
31	27	75.0	1841_RS16_BACTN	Q9rg15 bacteroides
32	27	75.0	1881_AMPM_METFE	P22624 methanother
33	27	75.0	1971_RL22_MYCTU	P95054 mycobacteri

34	27	75.0	2051_H1E_STRPU	P19375 strongyloce
35	27	75.0	2061_H1_ONCMY	P06350 oncorhynch
36	27	75.0	2081_DBH_MYCSM	O9zhc5 mycobacteri
37	27	75.0	2171_H1_ANAPL	P09426 anas platyr
38	27	75.0	2171_H1_CHICK	P09987 gallus galli
39	27	75.0	2181_H101_CHICK	P08284 gallus galli
40	27	75.0	2181_H11R_CHICK	P08284 gallus galli
41	27	75.0	2191_H110_CHICK	P08286 gallus galli
42	27	75.0	2191_H1B_XENLA	P06893 xenopus lae
43	27	75.0	2221_H15_MOUSE	P43276 mus musculu
44	27	75.0	2231_H103_CHICK	P08285 gallus galli
45	27	75.0	2241_H11L_CHICK	P08287 gallus galli

ALIGNMENTS

```

RESULT 1
EBP2_SCHPO STANDARD: PRT: 333 AA.
ID EBP2_SCHPO
AC O13802:
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable RNA processing protein EBP2 homolog.
GN SPAC17H9.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21846401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Woilckeart G., Aert R., Robben J., Grymoprez B.,
RA Welljans I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Egger P., Zimmermann W., Wedler H., Wambolt R., Punnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Motter S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Roohet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutis L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Snopcewski G.V., Ussery D., Barrett B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe." ;
RT Nature 415:871-880(2002).
-i- FUNCTION: Required for the processing of the 27S pre-rRNA (By
similarity).
-i- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
-i- SIMILARITY: BELONGS TO THE EBP2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC EMBL: 29857; CAB1214.1; -
 CC Hypothetical protein; Ribosome biogenesis; Nuclear protein;
 KW Coiled coil.
 FT DOMAIN 11 26 ALA/LYS-RICH.
 FT DOMAIN 212 240 COILED COIL (POTENTIAL).
 SQ SEQUENCE 333 AA; 37818 MW; 8EDD4B55B30711F CRC64;
 Query Match 88.9%; Score 32; DB 1; Length 333;
 Best Local Similarity 87.5%; Pred. No. 15;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ARKAKA 8
 Db 11 AOKKAKA 18
 RESULT 2
 ID AKAS HUMAN STANDARD; PRT; 427 AA.
 AC P24588;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE A-kinase anchor protein 5 (A-kinase anchor protein 79 kDa) (AKAP 79)
 DE (CAMP-dependent protein kinase regulatory subunit II high affinity
 binding protein) (H21).
 GN AKAP5 OR AKAP79.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.
 RC TISSUE=Thyroid;
 RX MEDLINE=92380978; PubMed=1512224;
 RA Card D.W., Stofko-Hahn R.E., Fraser I.D.C., Cone R.D., Scott J.D.;
 RT "Localization of the CAMP-dependent protein kinase to the
 postsynaptic densities by A-kinase anchoring proteins.
 RT Characterization of AKAP 79.";
 RL J. Biol. Chem. 267:16816-16823(1992).
 RN [2]
 RP SEQUENCE OF 332-427 FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=92129278; PubMed=1733921;
 RA Hirsch A.H., Glatz S.B., Li Y., You Y., Rubin C.S.;
 RT "Cloning and expression of an intron-less gene for AKAP 75, an anchor
 protein for the regulatory subunit of CAMP-dependent protein kinase
 II beta.";
 RL J. Biol. Chem. 267:2131-2134(1992).
 RN [1]
 RP FUNCTION: MAY ANCHOR THE PKA PROTEIN TO CYTOSKELETAL AND/OR
 ORGANELLE-ASSOCIATED PROTEINS, TARGETING THE SIGNAL CARRIED BY
 CAMP TO SPECIFIC INTRACELLULAR EFFECTORS. ASSOCIATION WITH TO THE
 BETA2-ADRENERGIC RECEPTOR (BETA2-AR) NOT ONLY REGULATES BETA2-AR
 SIGNALING PATHWAY, BUT ALSO THE ACTIVATION BY PKA BY SWITCHING OFF
 THE BETA2-AR SIGNALING CASCADE.
 CC -1- SUBUNIT: BINDING PROTEIN FOR DIMER OF THE RII-BETA REGULATORY
 SUBUNIT OF CAMP-DEPENDENT PROTEIN KINASE (PKC) AND ALSO FOR THE
 PROTEIN KINASE C (PKC) AND THE PHOSPHATASE CALCINEURIN (PP2B).
 CC EACH ENZYME IS INHIBITED WHEN BOUND TO THE ANCHORING PROTEIN. ALSO
 CC BINDS THE BETA2-ADRENERGIC RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH PARTICULATE FRACTIONS.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN THE CEREBRAL CORTEX AND THE
 POSTSYNAPTIC DENSITIES OF THE FOREBRAIN, AND TO A LESSER EXTENT
 CC IN ADRENAL MEDULLA, LUNG AND ANTERIOR PITUITARY.
 CC -1- DOMAIN: RII-ALPHA BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC
 HELIX, COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
 CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
 CC -1- MISCELLANEOUS: THE N-TERMINAL REGION, WHICH IS HIGHLY BASIC, IS
 CC REQUIRED FOR INTERACTION WITH CALMODULIN.
 CC -1- SIMILARITY: TO OTHER AKAP PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 CC EMBL: M90359; AAA58363.1; -
 CC PIR: A43453; A43453.
 CC Genew: HGNC:375; AKAP5.
 DR MIM: 604688; -
 DR InterPro: IPR001573; PKINA_anch.
 DR Calmodulin-binding. 103
 FT DOMAIN 73 170
 FT DOMAIN 1
 FT DOMAIN 392 405
 FT MUTAGEN 392 392
 FT MUTAGEN 396 396
 FT MUTAGEN 400 400
 FT MUTAGEN 405 405
 FT MUTAGEN 408 408
 FT CONFLICT 407 407
 SQ SEQUENCE 427 AA; 47072 MW; A3D08AE8D0521408 CRC64;
 Query Match 88.9%; Score 32; DB 1; Length 427;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 RKAKA 8
 Db 40 RKAKA 46
 RESULT 3
 ID AKAS BOVIN STANDARD; PRT; 428 AA.
 AC P24275;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE A-kinase anchor protein 5 (A-kinase anchor protein 75 kDa) (AKAP 75)
 DE (CAMP-dependent protein kinase regulatory subunit II high affinity
 binding protein) (P75).
 GN AKAP5 OR AKAP75.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=92129278; PubMed=1733921;
 RA Hirsch A.H., Glatz S.B., Li Y., You Y., Rubin C.S.;
 RT "Cloning and expression of an intron-less gene for AKAP 75, an anchor
 protein for the regulatory subunit of CAMP-dependent protein kinase
 II beta.";
 RL J. Biol. Chem. 267:2131-2134(1992).
 RN [2]
 RP SEQUENCE OF 170-428.
 RC TISSUE=Brain;
 RX MEDLINE=91201382; PubMed=2016323;
 RA Bregman D.B., Hirsch A.H., Rubin C.S.;
 RT "Molecular characterization of bovine brain P75, a high affinity
 binding protein for the regulatory subunit of CAMP-dependent protein
 kinase II beta.";
 RL J. Biol. Chem. 266:7207-7213(1991).
 RN [1]
 RP FUNCTION: MAY ANCHOR THE KINASE TO CYTOSKELETAL AND/OR ORGANELLE-
 ASSOCIATED PROTEINS, TARGETING THE SIGNAL CARRIED BY CAMP TO
 CC SPECIFIC INTRACELLULAR EFFECTORS.
 CC -1- SUBUNIT: BINDS DIMER OF THE RII-BETA REGULATORY SUBUNIT OF CAMP-
 CC DEPENDENT PROTEIN KINASE.

CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH PARTICULATE FRACTIONS.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN BRAIN, AND TO A LESSER EXTENT
CC IN ADRENAL MEDULLA, LUNG AND ANTERIOR PITUITARY.
CC -1- MISCELLANEOUS: THE N-TERMINAL REGION, WHICH IS HIGHLY BASIC, IS
CC REQUIRED FOR INTERACTION WITH CALMODULIN.
CC -1- SIMILARITY: TO OTHER AKAP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M82914: AAA30366.1; -;
DR EMBL: M60292: AAA30682.1; -;
DR PIR: A42311: A42311.
DR InterPro: IPR001573: PKinA_anch.
KW Phosphorylation; Brain; Calmodulin-binding.
FT DOMAIN 73 103
FT DOMAIN 1 170
FT ESSENTIAL TO THE INTRACELLULAR ANCHORING
FT FUNCTION.
FT DOMAIN 389 410
FT MOD_RES 87 87 RII-BETA SUBUNIT BINDING DOMAIN.
FT MOD_RES 92 92 PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT MOD_RES 94 94 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT MOD_RES 94 94 PHOSPHORYLATION (BY PKC) (POTENTIAL).
FT CONFLICT 225 225 R -> T (IN REF. 2).
SQ SEQUENCE 428 AA; 47878 MW; 427EBE980DB3173E CRC64;

Query Match 88.9%; Score 32; DB 1; Length 428;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RKKAKA 8
DB 40 RKKAKA 46
ID 1111111
ID Y617_SYNY3 STANDARD; PRT; 267 AA.
AC Q55707;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein sl10617.
GN SL10617.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_Taxid=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugitara M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- SIMILARITY: BELONGS TO THE PSPA/IM30 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D64002: BAA10345.1; -;
DR Hypothetical protein; Coiled coil; Complete proteome.
KW DOMAIN 26 156
FT COILED COIL (POTENTIAL).

SO SEQUENCE 267 AA; 28905 MW; 5DDE309FFB0FF1A6 CRC64;

Query Match 83.3%; Score 30; DB 1; Length 267;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKAKA 8
DB 141 ARKAKA 148
ID 1111111
ID H12_MOUSE STANDARD; PRT; 404 AA.
AC P57795;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable cysteine desulfurase (EC 4.4.1.-) (Nlfs protein homolog).
GN ISCS OR Nlfs.
OS Methanosarcina thermophila.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_Taxid=2210;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20389605; PubMed=10930739;
RA Borup B., Ferry J.G.;
RT "Cysteine biosynthesis in the archaea: Methanosarcina thermophila
RT utilizes O-acetylserine sulfhydrylase.";
RL FEWS Microbiol. Lett. 189:205-210(2000).
CC -1- FUNCTION: CATALYZES THE REMOVAL OF ELEMENTAL SULFUR FROM CYSTEINE
CC TO PRODUCE ALANINE (BY SIMILARITY).
CC -1- COPFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES. Nlfs/ISCS SUBFAMILY.
CC -1- CAUTION: The conserved pyridoxal-binding site Lys at position 216
CC is replaced by a Glu.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF276772: AAG01802.1; -;
DR InterPro: IPR000192: AminotransfV.
DR Pfam: PF00266; aminotran_5.1.
DR PROSITE: PS00595; AA_TRANSF_R_CLASS_5; FALSE_NEG.
KW Lyase; Pyridoxal phosphate. BY SIMILARITY.
FT ACT_SITE 338 338
SQ SEQUENCE 404 AA; 44335 MW; 87B5BA25F87E2A25 CRC64;

Query Match 83.3%; Score 30; DB 1; Length 404;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKAKA 8
DB 64 ARKAKA 71
ID 1111111
ID H12_MOUSE STANDARD; PRT; 211 AA.
AC P15864;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone H1.2 (H1 VAR.1) (H1C).
GN H1P2.

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69386668; PubMed=2780558;
RA Cheng G., Nandi A., Clerk S., Skoultschik A.I.;
RT "Different 3'-end processing produces two independently regulated
RT mRNAs from a single H1 histone gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7002-7006(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86059119; PubMed=2824517;
RA Yang Y.-S., Brown D.T., Wellman S.E., Sittman D.B.;
RT "Isolation and characterization of a mouse fully replication-dependent
RT H1 gene within a genomic cluster of core histone genes.";
RL J. Biol. Chem. 267:17118-17125(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=BALE/c;
RX MEDLINE=98322109; PubMed=9655912;
RA Franke K., Drahent B., Doecke D.;
RT "Expression of murine H1 histone genes during postnatal development.";
RL Biochim. Biophys. Acta 1398:232-242(1998).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M25365; AAA37808.1; -
DR EMBL; J03482; AAA37807.1; -
DR EMBL; Y12291; CAA72970.1; -
DR PIR; A28470; A28470.
DR PIR; A41389; A41389.
DR HSSP; P08287; 1GHC.
DR MGD; MGI:1931526; H1f2.
DR InterPro; IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone_1.
DR ProDom; PD000373; linker_hist_N_1.
DR SMART; SM00526; H1S; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation.
FT INIT_MET 0 0 ACETYLATION (BY SIMILARITY).
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT DOMAIN 35 113 GLOBULAR.
SQ SEQUENCE 211 AA; 21135 MW; F5C28D5F53279843 CRC64;

```

Query Match 80.6%; Score 29; DB 1; Length 211;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKA 7
1:|||||
DB 19 ARKKA 25

RESULT 7
SURF_HUMAN STANDARD; PRT; 361 AA.
AC 075683; Q9UK24; Q9BRK9; Q9BT25;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Surf6 locus protein 6.

```

GN SURF6 OR SURF-6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Magoules C., Fried M.;
RT "Isolation and genomic analysis of the human Surf-6 gene: a member of
RT the surf6 locus.";
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung, and Lymph;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RA Angiolillo A., Russo G., Porcellini A., Smaidone S., Scognamiglio B.,
RA D'Alessandro F., Pietropaolo C.T.;
RT "The human homologue of the mouse Surf-5 gene encodes multiple
RT alternatively spliced transcripts.";
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN A NUCLEOLAR FUNCTION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: BELONGS TO THE SURF6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF186772; AAD56587.1; -
DR EMBL; BC003001; AAH03001.1; -
DR EMBL; BC006197; AAH06197.1; -
DR EMBL; BC014878; AAH14878.1; -
DR EMBL; AJ224639; CAA12054.1; -
DR Genew; HGNC:11478; SURF6.
DR MIM; 185642; -
KW Nuclear protein.
FT DOMAIN 45 49 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 154 158 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 6 6 A -> T (IN REF. 3).
FT CONFLICT 102 102 D -> N (IN REF. 1).
SQ SEQUENCE 361 AA; 41450 MW; 34F63992E3E82797 CRC64;

```

Query Match 80.6%; Score 29; DB 1; Length 361;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 RKKAKA 8
|||||:|
DB 331 RKKARA 337

RESULT 8
ABF2_HUMAN STANDARD; PRT; 623 AA.
AC Q9UG63; O60864;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family F, member 2 (iron inhibited ABC
DE transporter 2) (HUS5Y-18).
GN ABCF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606; [1]

```

RP SEQUENCE FROM N.A.
RC TISSUE=uterus;
RA Ansoore W., Wirner U., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20403900; PubMed=10944468;
RA Ye 2., Connor J.R.;
RT "cDNA cloning by amplification of circularized first strand cDNAs
RL reveals non-IRE-regulated iron-responsive mRNAs.";
RL Blochem. Biophys. Res. Commun. 275:223-227(2000).
RN
RP [3]
RP SEQUENCE OF 74-623 FROM N.A.
RC TISSUE=lung;
RX MEDLINE=21064499; PubMed=11124703;
RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,
RA Cannata N., Zimbelli R., Lanfranchi G., Valle G.;
RT "Characterization of 16 novel human genes showing high similarity to
RT yeast sequences.";
RL Yeast 18:69-80(2001).
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ERF SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AL050291; CAB43392.1; -
DR EMBL; AF261091; AAG13902.1; -
DR EMBL; AJ005016; CAA06290.1; -
DR Genew; HGNC:71; ABCP2.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 2.
DR Prodom; PD000006; ABC_transportr; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport; Repeat.
FT NP_BIND 118 125 ATP (POTENTIAL).
FT NP_BIND 430 437 ATP (POTENTIAL).
FT CONFLICT 39 39 V -> A (IN REF. 2).
FT CONFLICT 74 74 T -> A (IN REF. 3).
FT CONFLICT 623 623 V -> VCHLTLASLPRP (IN REF. 1).
SO SEQUENCE 623 AA; 71290 MW; 702A968BCF8061AE CRC64;

Query Match 80.6%; Score 29; DB 1; Length 623;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKA 7
DB 6 ARKKA 12

RESULT 9
NOL1_HUMAN STANDARD; PRT; 855 AA.
AC P46087;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proliferating-cell nuclear antigen p120 (Proliferation-associated
DE nuclear protein p120).
GN NOL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RX MEDLINE=90315275; PubMed=2576976;
RA Fonagy A., Henning D., Jhlang S., Halder M., Busch R.K., Larson R.,

```

```

RA Valdez B., Busch H.;
RT "Cloning of the cDNA and sequence of the human proliferating-cell
RT nuclear protein p120.";
RL Cancer Commun. 1:243-251(1989).
RN
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90321785; PubMed=2372471;
RA Larson R.G., Henning D., Halder M.A., Jhlang S., Lin W.L., Zhang W.W.,
RA Busch H.;
RT "Genomic structure of the human proliferating cell nuclear protein
RT p120.";
RL Cancer Commun. 2:63-71(1990).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE REGULATION OF THE CELL CYCLE AND
CC THE INCREASED NUCLEOLAR ACTIVITY THAT IS ASSOCIATED WITH THE CELL
CC PROLIFERATION.
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN G1 AND PEAKS DURING THE EARLY S
CC PHASE OF THE CELL CYCLE.
CC -1- SIMILARITY: BELONGS TO THE SUN (BACTERIAL) / NUCLEOLAR PROTEIN
CC NOL1/NOP2 (EUKARYOTES) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M32110; AAA36398.1; ALT_SEQ.
DR EMBL; M33132; -; NOT_ANNOTATED_CDS.
DR EMBL; X55504; CAA39119.1; -
DR PIR; A48168; A48168.
DR SWISS-2DPAGE; P46087; HUMAN.
DR Genew; HGNC:7867; NOL1.
DR MIM; 164031; -
DR InterPro; IPR001678; Sun_Nop1/Nop2.
DR Pfam; PF01189; NOL1_Nop2_Sun; 1.
DR TIGRFSMS; TIGR00446; nop2p; 1.
DR PROSITE; PS01153; NOL1_NOP2_SUN; 1.
KW Nuclear protein; Antigen.
SO SEQUENCE 855 AA; 94078 MW; 43DC4ECB98A5862 CRC64;

Query Match 80.6%; Score 29; DB 1; Length 855;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKA 7
DB 47 ARKKA 53

RESULT 10
XPC_MOUSE STANDARD; PRT; 900 AA.
AC P51612; P54732;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE DNA-repair protein complementing XP-C cells homolog (Xeroderma
DE pigmentosum group C complementing protein homolog) (P125).
GN XPC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RX MEDLINE=96184849; PubMed=8604333;
RA Li L., Peterson C., Legeurski R.;
RT "Sequence of the mouse XPC cDNA and genomic structure of the human
RT XPC gene.";
RL Nucleic Acids Res. 24:1026-1028(1996).

```

RN [2]
 RP SEQUENCE OF 28-587 FROM N.A.
 RC STRAIN-129/SV;
 RA MEDLINE=95405469; PubMed=7675084;
 RA Sands A.T., Apolin A., Sanchez A., Conli C.J., Bradley A.;
 RT "High susceptibility to ultraviolet-induced carcinogenesis in mice
 RT lacking xpc.";
 RL Nature 377:162-165(1995).
 CC -i- FUNCTION: INVOLVED IN DNA EXCISION REPAIR. MAY PLAY A PART IN DNA
 CC DAMAGE RECOGNITION AND/OR IN ALTERING CHROMATIN STRUCTURE TO
 CC ALLOW ACCESS BY DAMAGE-PROCESSING ENZYMES.
 CC -i- SUBUNIT: HETERODIMER OF A 125 kDa SUBUNIT (P125) AND OF A
 CC 58 kDa SUBUNIT (P58).
 CC -i- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -i- SIMILARITY: SOME, TO YEAST RAD4.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U27398; AAC52500.1; -
 DR EMBL: U40005; AAA82720.1; -
 DR MGD: MGI:103557; Xpc.
 DR InterPro: IPR004583; Rad4.
 DR TIGRFAMS: TIGR00605; Rad4; 1.
 KW DNA repair; DNA-binding; Nuclear protein.
 FT CONFLICT 28 30 AVN -> CSD (IN REF. 2).
 FT CONFLICT 53 53 S -> L (IN REF. 2).
 FT CONFLICT 67 67 L -> F (IN REF. 2).
 FT CONFLICT 70 70 L -> S (IN REF. 2).
 FT CONFLICT 134 135 RG -> TP (IN REF. 2).
 FT CONFLICT 165 170 EVOENM -> GVHEHD (IN REF. 2).
 FT CONFLICT 181 181 S -> N (IN REF. 2).
 FT CONFLICT 187 187 S -> N (IN REF. 2).
 FT CONFLICT 190 190 R -> S (IN REF. 2).
 FT CONFLICT 192 192 P -> L (IN REF. 2).
 FT CONFLICT 342 345 GSKA -> AKP (IN REF. 2).
 FT CONFLICT 367 367 R -> S (IN REF. 2).
 FT CONFLICT 428 428 R -> C (IN REF. 2).
 FT CONFLICT 467 467 C -> S (IN REF. 2).
 FT CONFLICT 584 584 K -> E (IN REF. 2).
 SQ SEQUENCE 900 AA; 100873 MW; 95737FCB36DC15DD CRC64;
 Query Match 80.6%; Score 29; DB 1; Length 900;
 Best local Similarity 75.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kuyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
 RA Raymond C., Rouse G., Seshimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dotan M.,
 RA Chunley F., Tingey S.V., Tomb J.F., Gordon M.P., Olson M.V.,
 RA Nestor E.W.;
 RT "The genome of the natural genetic engineer *Agrobacterium tumefaciens*
 RT C58.";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Guroilo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houniel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT *Agrobacterium tumefaciens* C58.";
 RL Science 294:2323-2328(2001).
 CC -i- SIMILARITY: BELONGS TO THE L29P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AE009147; AAL42934.1; -
 DR EMBL: AE008112; AAK87700.1; -
 DR PROSITE: PS00579; RIBOSOMAL_L29; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 66 AA; 7485 MW; 340B2BF34D4F99A CRC64;
 Query Match 77.8%; Score 28; DB 1; Length 66;
 Best local Similarity 75.0%; Pred. No. 22;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- SIMILARITY: BELONGS TO THE L29P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL551787; CAC45943.1; -
DR InterPro: IPR001854; Ribosomal_L29.
DR Pfam: PF00831; Ribosomal_L29; 1.
DR TIGRfams: TIGR00012; L29; 1.
DR PROSITE: PS00579; RIBOSOMAL_L29; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 66 AA; 7483 MW; AED19B44E5072483 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 66;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKRAKA 8
Db 57 ARORAAEA 64
|||||:
RS13_SULSO STANDARD; PRT; 165 AA.
AC P95986;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S13p/S18E.
GN RPS13P OR RPS13AB OR RPS13 OR SSC0074 OR CO4_048.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=97055432; PubMed=8899719;
RA Sersen C.W., Kleink H.-P., Singh R.K., Allard G., Chan C.C.-Y.,
RA Liu Q.-Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
RA Doolittle W.F., Ragan M.A., Charlebois R.L.;
RT "Organizational characteristics and information content of an
RT archaeal genome: 156 kb of sequence from Sulfolobus solfataricus
RT P2."
RL Mol. Microbiol. 22:175-191(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayer M.J., Chan-Weiler C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sersen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -1- FUNCTION: INVOLVED IN THE BINDING OF EMET-TRNA AND, HENCE, IN THE
CC INITIATION OF TRANSLATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y08257; CAA69528.1; -
DR EMBL: AE006647; AAK40436.1; -
DR InterPro: IPR001892; Ribosomal_S13.
DR Pfam: PF00416; Ribosomal_S13; 1.
DR PRODOM: PD001363; Ribosomal_S13; 1.
DR PROSITE: PS00646; RIBOSOMAL_S13; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 165 AA; 18629 MW; 8F4DAABEB8736751 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 165;
Best Local Similarity 85.7%; Pred. No. 51;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKRAAK 7
Db 145 ARKRAAQ 151
|||||:
RK6_GUTTH STANDARD; PRT; 179 AA.
AC O46908;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Chloroplast 50S ribosomal protein L6.
GN RPL6.
OS Guillardia theta (Cryptomonas phl).
OC Chloroplast.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97283757; PubMed=9137835;
RA Wang S.L., Liu X.-Q., Douglas S.E.;
RT "The large ribosomal protein gene cluster of a cryptomonad plastid:
RT gene organization, sequence and evolutionary implications."
RT Biochem. Mol. Biol. Int. 41:1035-1044(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=99128221; PubMed=9929392;
RA Douglas S.E., Penny S.L.;
RT "The plastid genome of the cryptophyte alga, Guillardia theta:
RT complete sequence and conserved synteny groups confirm its common
RT ancestry with red algae."
RL J. Mol. Evol. 48:236-244(1999).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
CC LOCATED AT THE AMINOACYL-TRNA BINDING SITE OF THE
CC PEPTIDYLTRANSFERASE CENTER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L6P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF041468; AAC35717.1; -
DR HSSP: P02391; 1RL6.
DR InterPro: IPR000702; Ribosomal_L6.
DR InterPro: IPR002358; Ribosomal_L6_1.
DR Pfam: PF00347; Ribosomal_L6; 1.
DR PRINTS: PR00059; RIBOSOMAL_L6.
DR PRODOM: PD002236; Ribosomal_L6; 1.
DR PROSITE: PS00525; RIBOSOMAL_L6_1; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 179 AA; 19527 MW; 8B4C0BDB0152AD24 CRC64;

```

Query Match 77.8%; Score 28; DB 1; Length 179;

Best Local Similarity 85.7%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 RKKAKA 8
DB 170 RKKAGKA 176

Search completed: April 8, 2003, 11:24:38
Job time : 8.2069 secs

RESULT 15

SODM_ASPFU STANDARD; PRT; 210 AA.

AC Q92450; 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Mn], mitochondrial precursor (EC 1.15.1.1)
DE (Allergen Asp f 6).
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 4202 / AF-102;
RX MEDLINE=96305209; PubMed=8691141;
RA Cramerl R., Falth A., Hemmann S., Jaussi R., Ismail C., Menz G.,
RA Blaser K.;
RT "Humoral and cell-mediated autoimmunity in allergy to Aspergillus
fumigatus.";
RL J. Exp. Med. 184:265-270(1996).
CC -I- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -I- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -I- COFACTOR: Manganese (By similarity).
CC -I- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -I- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U53561; AAB60779.1; ALT_INIT.
DR HSSP: P04179; IABM.
DR InterPro: IPR001189; SODismutase.
DR Pfam: PF00081; sodfe_1.
DR Pfam: PF02777; sodfe_C; 1.
DR ProDom: PD000475; SODismutase; 1.
DR PROSITE: PS00088; SOD_MN; 1.
KW Oxidoreductase; Manganese; Mitochondrion; Transit peptide; Allergen.
FT TRANSIT 1 ? MITOCHONDRION (BY SIMILARITY).
FT CHAIN ? 210 SUPEROXIDE DISMUTASE [MN].
FT METAL 29 29 MANGANESE (BY SIMILARITY).
FT METAL 77 77 MANGANESE (BY SIMILARITY).
FT METAL 163 163 MANGANESE (BY SIMILARITY).
FT METAL 167 167 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 210 AA: 23377 MW: CEG4A134780E5546 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 210;

Best Local Similarity 75.0%; Pred. No. 64;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKA 8
DB 47 AKKAA 54

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:17:38 ; Search time 28.4138 seconds
(without alignments)
58.013 Million cell updates/sec

Title: US-09-496-391-5
Perfect score: 36
Sequence: 1 ARKKAKA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	91.7	187	3	09P3T4
2	33	91.7	473	5	017155
3	33	91.7	1062	5	09G2U0
4	32	88.9	369	5	0961L5
5	32	88.9	478	10	09STW1
6	32	88.9	487	5	095ZV8
7	32	88.9	537	5	08T8V7
8	32	88.9	564	5	09VRX8
9	32	88.9	2045	16	09A0K5
10	31	86.1	105	17	029591
11	30	83.3	166	16	08Y6S1
12	30	83.3	218	10	08R2R9
13	30	83.3	232	10	039576
14	30	83.3	300	16	09KY91
15	30	83.3	308	10	08R0H1
16	30	83.3	438	2	09E2J8

17	30	83.3	485	10	08RXD0	08rx0 arabidopsis
18	30	83.3	781	17	09P9N1	09p9n1 sulforispha
19	30	83.3	781	17	096YV1	096yv1 sulfolobus
20	30	83.3	790	17	08ZS3	08zs3 pyrobaculum
21	30	83.3	924	10	09S0U8	09s0u8 arabidopsis
22	29	80.6	97	16	08U537	08u537 agrobacteri
23	29	80.6	144	13	092167	092167 xenopus lae
24	29	80.6	185	16	08XXW2	08xxw2 ralsionia s
25	29	80.6	191	5	046140	046140 mytilus edu
26	29	80.6	191	5	046141	046141 mytilus edu
27	29	80.6	191	5	046142	046142 mytilus edu
28	29	80.6	191	5	046143	046143 mytilus edu
29	29	80.6	191	5	046362	046362 mytilus edu
30	29	80.6	227	16	09JY90	09jy90 neisseria m
31	29	80.6	262	10	09LH49	09lh49 arabidopsis
32	29	80.6	277	16	09XA03	09xa3 streptomyce
33	29	80.6	298	2	0910U5	091u5 rhizobium m
34	29	80.6	309	5	019681	019681 caenorhabdi
35	29	80.6	324	10	023385	023385 arabidopsis
36	29	80.6	352	16	09XA14	09xa14 streptomyce
37	29	80.6	374	10	08M4H9	08m4h9 arabidopsis
38	29	80.6	378	5	P91080	P91080 caenorhabdi
39	29	80.6	394	5	095QJ3	095qj3 caenorhabdi
40	29	80.6	395	2	09LAV2	09lav2 streptococc
41	29	80.6	442	2	09F7P1	09f7p1 uncultured
42	29	80.6	448	16	09HVK6	09hvk6 pseudomonas
43	29	80.6	461	16	092MX2	092mx2 rhizobium m
44	29	80.6	464	16	08RC72	08rc72 thermomane
45	29	80.6	469	17	08TJ37	08tj37 methanosarc

ALIGNMENTS

RESULT 1

09P3T4 PRELIMINARY; PRT; 187 AA.
ID 09P3T4
AC 09P3T4;
DT 01-OCR-2000 (TREMBlrel. 15, Created)
DT 01-OCR-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Conserved hypothetical protein.
GN B24P7.10.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL389890; CAB97266.1; -;
DR InterPro: IPR001950; TIF_SUI1.
DR Pfam: PF01253; SUI1; 1.
DR TIGRFAMs: TIGR01159; DRP1; 1.
DR PROSITE: PS0296; SUI1_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 187 AA; 20756 MW; 27F5AD68D6127EAB CRC64;

Query Match 91.7%; Score 33; DB 3; Length 187;
Best Local Similarity 87.5%; Pred. No. 33;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ARKKAKA 8
Db 71 ARKKAKA 78

```

RESULT 2
ID 017155 PRELIMINARY: PRT: 473 AA.
AC 017155:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE C41H7.7 protein.
GN C41H7.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Barks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Geisel C., Mamsley P., Elliott G., Smith A.;
RT "The sequence of C. elegans cosmid C41H7."
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR EMBL: AF025450; AAB70938.1; -.
DR HSP: P06734; 1H1F.
DR InterPro: IPR000859; CUB_domain.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00431; CUB; 1.
DR Pfam: PF00059; lectin_c; 2.
DR SMART: SM00034; CLECT; 2.
DR SMART: SM00042; CUB; 1.
DR PROSITE: PS01180; CUB; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; UNKNOWN_1.
DR PROSITE: PS00041; C_TYPE_LLECTIN_2; 2.
SQ SEQUENCE 473 AA; 52600 MW; 23373070152C0C09 CRC64;

Query Match 91.7%; Score 33; DB 5; Length 473;
Best Local Similarity 87.5%; Pred. No. 76;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKAKA 8
Db 199 SRKAKA 206

RESULT 3
O9GZJ0 PRELIMINARY: PRT: 1062 AA.
AC 09GZJ0:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE L8149.6.
GN L8149.6.

```

```

OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxId=5664;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Myler P.J., Sisk E., Cawthra J., Handley F., Vogt C., Robertson L.,
RA McDonagh P., Stuart K., Ivens A., Worthey E.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC016161; AAF98397.1; -.
SQ SEQUENCE 1062 AA; 115906 MW; 9563FB3C686C56A4 CRC64;

Query Match 91.7%; Score 33; DB 5; Length 1062;
Best Local Similarity 87.5%; Pred. No. 16+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKAKA 8
Db 210 ARKAKA 217

RESULT 4
O961L5 PRELIMINARY: PRT: 369 AA.
AC 0961L5:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE GH17930P.
GN SNS OR CG2385 OR CG8278 OR CG12495 OR CG13752 OR CG13753 OR CG13754
OR CG13755 OR CG18464.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY051522; AAK92946.1; -.
DR FlyBase: FBgn0024189; sns.
DR InterPro: IPR001230; Prenyl_site.
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
SQ SEQUENCE 369 AA; 41753 MW; 676CE63BB03F465D CRC64;

Query Match 88.9%; Score 32; DB 5; Length 369;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 RKKAKA 8
Db 33 RKKAKA 39

RESULT 5
O9STW1 PRELIMINARY: PRT: 478 AA.
AC 09STW1:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE Hypothetical 53.9 kDa protein.
GN T22A6.160 OR AT4G24330.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

```


OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Beyer M., Zimmermann W., Grueneisen A., Wambutt R., Bancroft I.,
RL Mewes H.W., Mayer K.F.X., Lemcke K., Schuelter C.,
RN Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Grueneisen A., Wambutt R., Kalicki J., Wohldmann P.,
RL Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.,
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL078637; CAB45068.1; -
DR EMBL: AL161561; CAB79343.1; -
KW Hypothetical protein.
SQ SEQUENCE 478 AA; 53947 MW; 896D16A5B907B55 CRC64;

Query Match 88.9%; Score 32; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKRAK 7
Db 278 ARKRAK 284
|||||
Q952V8 PRELIMINARY; PRT; 487 AA.
ID Q952V8
AC Q952V8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 56.3 kDa protein.
GN F08F8.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du Z.;
RT "The sequence of C. elegans cosmid F08F8.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: U28991; AAK68306.2; -
KW Hypothetical protein.
SQ SEQUENCE 487 AA; 56350 MW; A12D6A7B25BCA6A6A CRC64;

Query Match 88.9%; Score 32; DB 5; Length 487;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARKRAK 7
Db 166 ARKRAK 172
|||||
Q8T8V7 PRELIMINARY; PRT; 537 AA.
ID Q8T8V7
AC Q8T8V7;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE AT21853P.
GN CG8398.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Paclebo J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Cejner S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY075255; ALU68122.1; -
SQ SEQUENCE 537 AA; 61098 MW; CEF0071D8B8F48CD CRC64;

Query Match 88.9%; Score 32; DB 5; Length 537;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARKRAK 8
Db 296 ARKRAK 303
|||||
Q9VRX8 PRELIMINARY; PRT; 564 AA.
ID Q9VRX8
AC Q9VRX8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG8398 protein.
GN CG8398.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Karlush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden Klamos I., Simpson M., Skupsky M.P., Smith T.,
RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195 (2000).
DR EMBL: AE003562; AAF50653.1; -
DR Flybase: FBgn0035708; CG8398.
DR InterPro: IPR002203; InterIn.
DR PROSITE: PS00881; PROTEIN_SPLICING; UNKNOWN.1.
SQ SEQUENCE 564 AA; 63927 MW; 31C078D9211DA84 CRC64;

Query Match 88.9%; Score 32; DB 5; Length 564;
Best Local Similarity 87.5%; Pred. No. 14e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKAKA 8
Db 296 ARKAKA 303

RESULT 9
O9A0K5 PRELIMINARY; PRT; 2045 AA.
AC O9A0K5:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative extracellular matrix binding protein.
GN EPR OR SPY0737.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SPROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
DR EMBL: AE006525; AAK33683.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR TIGRFAMS: TIGR01167; LPTXG_anchor.1.
KM Complete proteome.
SQ SEQUENCE 2045 AA; 221960 MW; 22C266867F53F19B CRC64;

Query Match 88.9%; Score 32; DB 16; Length 2045;
Best Local Similarity 87.5%; Pred. No. 4e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKAKA 8

Db 1671 ARKAKA 1678

RESULT 10
O29591 PRELIMINARY; PRT; 105 AA.
AC O29591:
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein AF0666.
GN AF0666.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerecavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodok A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utermack T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon *Archaeoglobus fulgidus*.";
RL Nature 390:364-370 (1997).
DR EMBL: AE001059; AAB90575.1; -
DR TIGR: AF0666; -
DR InterPro: IPR002852; DUF134.
DR Pfam: PF02001; DUF134; 1.
DR Prodom: PD015339; DUF134; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 105 AA; 11691 MW; 8BBAD103BFCFAF61 CRC64;

Query Match 86.1%; Score 31; DB 17; Length 105;
Best Local Similarity 87.5%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKAKA 8
Db 80 ARKAKA 87

RESULT 11
O8Y651 PRELIMINARY; PRT; 166 AA.
AC O8Y651:
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein BME11311.
GN BME11311.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA Delvecchio V.G., Kapartel V., Redkar R.J., Patra G., Muier C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jadhavani L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Teleson J.-J.,

RA Haselkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT *Brucella melitensis*."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009569; AL52492.1; -.
DR InterPro: IPR00104; Antifreeze_1.
DR PRINTS: PR00308; ANTIFREEZE1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 166 AA; 17084 MW; EB0C097797ABE856 CRC64;

Query Match 83.3%; Score 30; DB 16; Length 166;
Best Local Similarity 87.5%; Pred. NO. 1.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKKAARA 8
DB 28 ARKKAARA 35

RESULT 12

08RZR9 PRELIMINARY; PRT; 218 AA.
AC 08RZR9;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE B1140D12.10 protein.
GN B1140D12.10.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa niponbare(GA3) genomic DNA, chromosome 1, BAC
RT clone:B1140D12.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003537; BAB86162.1; -.
SQ SEQUENCE 218 AA; 23468 MW; 9B158986E7938F29 CRC64;

Query Match 83.3%; Score 30; DB 10; Length 218;
Best Local Similarity 87.5%; Pred. NO. 1.6e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKKAARA 8
DB 18 ARKKAARA 25

RESULT 13

039576 PRELIMINARY; PRT; 232 AA.
AC 039576;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Histone H1.
GN CH1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120862; PubMed=8590479;
RA Fabry S., Muller K., Lindner A., Park P.B., Cornelius T., Schmitt R.;
RT "The organization structure and regulatory elements of Chlamydomonas
RT histone genes reveal features linking plant and animal genes."
RL Curr. Genet. 28:333-345(1995).
DR EMBL: U16726; AAA98452.1; -.

DR HSSP: P02259; 1HST.
DR InterPro: IPR001386; Histone_H1/H5.
DR InterPro: IPR003216; Linkerhist.N.
DR Pfam: PF00538; linker_histone_1.
DR ProDom: PD000373; linkerhist_LN; 1.
DR SMART: SM00526; H15; 1.
SQ SEQUENCE 232 AA; 24693 MW; 2D006AE448AFA037 CRC64;

Query Match 83.3%; Score 30; DB 10; Length 232;
Best Local Similarity 87.5%; Pred. NO. 1.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARKKAARA 8
DB 14 ARKKAARA 21

RESULT 14

09KY91 PRELIMINARY; PRT; 300 AA.
AC 09KY91;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative integral membrane protein.
GN SC05011 OR SCK15.13.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford R., Ruter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL: AL356813; CAB92604.1; -.
SQ SEQUENCE 300 AA; 32763 MW; 3BD1EA5DD2BB7F4 CRC64;

Query Match 83.3%; Score 30; DB 16; Length 300;
Best Local Similarity 75.0%; Pred. NO. 2.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARKKAARA 8
DB 255 ARKKAARA 262

RESULT 15

08RUH1 PRELIMINARY; PRT; 308 AA.
AC 08RUH1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative 60S ribosomal protein.
GN B1033B05.7 OR P0592605.26.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.

```
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, BAC
RL clone:BI033B05.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
RL clone:P0592G05.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP004223; BAB90657.1; -.
DR EMBL: AP004672; BAB90840.1; -.
KW Ribosomal protein.
SQ SEQUENCE 308 AA; 34272 MW; C9BF0DE3B14D313C CRC64;

Query Match      83.3%; Score 30; DB 10; Length 308;
Best Local Similarity 87.5%; Pred.No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ARKKAKA 8
   |||||
Db 78 ARKKAAA 85
```

Search completed: April 8, 2003, 11:28:15
Job time : 32.4138 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:12:02 ; Search time 4.65517 Seconds
(without alignments)
53.458 Million cell updates/sec

Title: US-09-496-391-4
Perfect score: 27
Sequence: 1 AKKARA 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	27	100.0	210 1	H1_LYTPI
2	27	100.0	347 1	ITP2_HUMAN
3	27	100.0	651 1	KEI3_YEAST
4	27	100.0	1431 1	DP2L_PYRO
5	27	100.0	1455 1	DP2L_PYRAB
6	27	100.0	1455 1	HCT1_CHIMU
7	24	88.9	125 1	HCT1_CHIMU
8	24	88.9	135 1	H2A_TRYCR
9	24	88.9	146 1	CLSP_HUMAN
10	24	88.9	171 1	H1_ECHCR
11	24	88.9	176 1	YLX9_CAEEL
12	24	88.9	191 1	RR4_CYAME
13	24	88.9	204 1	Y981_METJA
14	24	88.9	218 1	YPOC_BACSU
15	24	88.9	235 1	H47_STEPL
16	24	88.9	248 1	H1_PARAN
17	24	88.9	250 1	RL7B_SCHPO
18	24	88.9	259 1	ALDC_KLETE
19	24	88.9	296 1	YEMA_SCHPO
20	24	88.9	298 1	CYPE_MOUSE
21	24	88.9	301 1	CYPE_HUMAN
22	24	88.9	428 1	G6PI_MYCGA
23	24	88.9	600 1	KU70_RHIAF
24	24	88.9	717 1	SKI_XENLA
25	24	88.9	727 1	VP4_RDVA
26	24	88.9	727 1	VP4_RDVF
27	24	88.9	1043 1	P11D_MOUSE
28	24	88.9	1044 1	P11D_HUMAN
29	24	88.9	1092 1	DP2L_METTH
30	24	88.9	1263 1	DP2L_PYRTH
31	23	85.2	80 1	R13E_AERPE
32	23	85.2	90 1	YJBD_ECOLI
33	23	85.2	93 1	SASG_BACCE

34	23	85.2	99 1	VHSB_BPT7	P03751 bacterioph
35	23	85.2	106 1	VHSB_BPT3	P20322 bacterioph
36	23	85.2	116 1	RL19_FUSNN	O8r880 fuscobacteri
37	23	85.2	120 1	RL35_YEAST	B93741 saccharomyc
38	23	85.2	121 1	RNPA_NEIMA	O9jw46 neisseria m
39	23	85.2	121 1	RNPA_NEIMA	O9jx56 neisseria m
40	23	85.2	126 1	SMD1_CAEEL	Q10013 caenorhabdi
41	23	85.2	135 1	ATPE_EUGGR	P11477 euglena gra
42	23	85.2	135 1	ATPE_SYNY3	P25533 synchoyst
43	23	85.2	136 1	ATPE_PRODI	P50010 prochloron
44	23	85.2	138 1	STP4_PIG	O09821 sus scrofa
45	23	85.2	148 1	H1L_MYTCA	P22974 mytilus cal

ALIGNMENTS

RESULT 1	ID	STANDARD	PRT	210 AA.
H1_LYTPI				
AC	P06144:			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JAN-1988 (Rel. 06, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Late histone H1.			
OS	Lytechinus pictus (Painted sea urchin).			
OC	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;			
OC	Echinoidea; Euechinoidea; Echinacea; Temnopneustidae;			
OC	Lytechinus.			
OX	NCBI_TaxID=7653;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Embryo;			
RX	MEDLINE=87040778; PubMed=3022245;			
RA	Knowles J.A., Childs G.J.;			
RT	"Comparison of the late H1 histone genes of the sea urchins			
RL	Lytechinus pictus and Strongylocentrotus purpuratus.";			
CC	Nucleic Acids Res. 14:8121-8133(1986).			
CC	-1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF			
CC	NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.			
CC	-1- SUBCELLULAR LOCATION: Nuclear.			
CC	-1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X04488; CA28177.1; -.			
DR	PIR: A25550; A25550.			
DR	HSSP: P02259; IHST.			
DR	InterPro: IPR001386; Histone_H1/H5.			
DR	InterPro: IPR003216; Linkerhist_N.			
DR	Pfam: PF00538; linker histone; 1.			
DR	Prodom: PD000373; linkerhist_N; 1.			
DR	SMART: SM00526; H15; 1.			
KW	Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.			
SQ	SEQUENCE 210 AA; 21746 MW; 08C38F6494007E2 CRC64;			
Query Match	100.0%; Score 27; DB 1; Length 210;			
Best Local Similarity	100.0%; Pred. No. 18;			
Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
OY	1 AKKARA 6			
Db	9 AKKARA 14			
RESULT 2				
ITP2_HUMAN				

```

ID      ITP2_HUMAN      STANDARD:      PRT:      347 AA.
AC      Q9URP3;
DT      15-JUN-2002 (Rel. 41, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Integrin beta 1 binding protein 2 (Melusin) (MSTP015).
GN      ITGB1BP2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE-Skeletal muscle;
RX      MEDLINE=9436136; PubMed=10506186;
RA      Brancaccio M., Guazzone S., Menini N., Sibona E., Hirsch E.,
RA      De Andrea M., Rocchi M., Altuda F., Tarone G., Silengo L.;
RT      "Melusin is a new muscle-specific interactor for beta(1) integrin
RT      cytoplasmic domain.";
RL      J. Biol. Chem. 274:29282-29288(1999).
CC      - FUNCTION: MAY PLAY A ROLE DURING MATURATION AND/OR ORGANIZATION OF
CC      MUSCLES CELLS.
CC      - SUBUNIT: INTERACTS WITH BETA-1 INTEGRIN SUBUNIT. THIS INTERACTION
CC      IS REGULATED BY DIVALENT CATIONS, AND IT OCCURS ONLY IN ABSENCE OF
CC      CALCIUM.
CC      - TISSUE SPECIFICITY: EXPRESSED IN SKELETAL AND CARDIAC MUSCLES BUT
CC      NOT IN OTHER TISSUES.
CC      - DOMAIN: THE TAIL DOMAIN BINDS TO THE CYTOPLASMIC DOMAIN OF BOTH
CC      INTEGRIN BETA-1A AND BETA-1D ISOFORMS. THE PRESENCE OF CA2+ IONS
CC      DOES NOT PREVENT BINDING OF A FRAGMENT CONSISTING OF THE SECOND
CC      CYSTEINE RICH REPEAT AND THE TAIL DOMAIN BUT PREVENTS THE BINDING
CC      OF THE FULL-LENGTH PROTEIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AF140690; AAF01676.1; -
DR      Genew: HGNC:6154; ITGB1BP2.
DR      MIM: 300332; -.
KW      SH3-binding.
FT      DOMAIN 5 59 CYS-RICH.
FT      SITE 149 203 CYS-RICH.
FT      SITE 28 31 SH3-BINDING (POTENTIAL).
FT      SITE 70 78 SH3-BINDING (POTENTIAL).
FT      SITE 172 175 SH3-BINDING (POTENTIAL).
FT      SITE 158 161 SH2-BINDING (POTENTIAL).
FT      SITE 234 237 SH2-BINDING (POTENTIAL).
FT      DOMAIN 320 347 ASP/GLU-RICH (ACIDIC).
SO      SEQUENCE 347 AA; 38382 MW; 1D0A94931821E74D CRC64;

Query Match 100.0%; Score 27; DB 1; Length 347;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AKKARA 6
DB      310 AKKARA 315

```

```

OS      Saccharomyces cerevisiae (baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-S288C / AB972;
RX      MEDLINE=9731371; PubMed=9169875;
RA      Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA      Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA      Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA      Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA      Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA      Duncan M., Floeth M., Fortin N., Friesen J.D., Fritze C., Goffeau A.,
RA      Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA      Hunkeler-Smith S., Hyman R., Johnston M., Kallman S., Kleene K.,
RA      Komp C., Kurd O., Laskaril D., Lew H., Lin A., Lin D., Louis E.J.,
RA      Marthe R., Messing F., Mewes H.-W., Mitrapati S., Moestl D.,
RA      Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA      Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA      Scherrens B., Schramm S., Schroeder W., Solcu A.M., Tettelin H.,
RA      Urestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA      Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA      Zhong W.W., Zollner A., Vo D.H., Han J.;
RT      "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL      Nature 387:103-105(1997).
CC      - SIMILARITY: CONTAINS 4 KELCH REPEATS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: 273619; CA97998.1; -
DR      SGD: S0006184; KEL3.
DR      InterPro: IPR001798; Kelch.
DR      Pfam: PF01344; Kelch; 3.
KW      Repeat.
FT      REPEAT 94 148 KELCH 1.
FT      REPEAT 150 199 KELCH 2.
FT      REPEAT 208 259 KELCH 3.
FT      REPEAT 262 310 KELCH 4.
SO      SEQUENCE 651 AA; 75433 MW; FD0C65A6AB30843 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 651;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AKKARA 6
DB      11 AKKARA 16

```

```

RESULT 3
ID      KEL3_YEAST      STANDARD:      PRT:      651 AA.
AC      Q08979;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Kelch repeats protein 3.
GN      KEL3 OR YPL263C.

```

```

RESULT 4
ID      DP2L_PYRHO      STANDARD:      PRT:      1431 AA.
AC      Q57861;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      DNA polymerase II large subunit (EC 2.7.7.7) (Pol II) [contains: Pho
DE      polc intein (Pho pol II intein)].
GN      POIC OR PHO121.
OS      Pyrococcus horikoshii.
OC      Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC      Pyrococcus.
OX      NCBI_TaxID=53953;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-OT3;

```



```

OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=83360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MOPn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utechtack T., Berry K., Bass S.,
RA Liner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Ewin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: MIGHT HAVE A ROLE ANALOGOUS TO THAT OF EUKARYOTIC
CC HISTONE PROTEINS (BY SIMILARITY).
CC -1- DEVELOPMENTAL STAGE: SPECIFIC TO THE EB (ELEMENTARY BODY) FORM
CC IN THE LIFE CYCLE OF CHLAMYDIAE.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY. HCT SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002279; AAF38997.1; -.
DR TIGR: TC0119; -.
KW DNA-binding; Repeat; Complete proteome.
SQ SEQUENCE 125 AA; 13683 MW; 5D2EC0E1CDE2P07 CRC64;

Query Match 88.9%; Score 24; DB 1; Length 125;
Best Local Similarity 83.3%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
Db 69 AKKARA 74

RESULT 7
HCT1_CHLTR
ID HCT1_CHLTR STANDARD: PRT; 125 AA.
AC Q02281; O84748;
DT 01-JUL-1993 (Rel. 26, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone H1-like protein HCL.
GN HCTA OR CT743.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=813;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-22.
RC STRAIN=L2/434/Bu;
RX MEDLINE=91210171; PubMed=1708378;
RA Tao S., Kaul R., Wenman W.M.;
RT "Identification and nucleotide sequence of a developmentally
RT regulated gene encoding a eukaryotic histone H1-like protein from
RT Chlamydia trachomatis."
RL J. Bacteriol. 173:2818-2822(1991).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=L2/434/Bu;
RX MEDLINE=91219487; PubMed=2023942;
RA Hackstadt T., Baehr W., Ying Y.;
RT "Chlamydia trachomatis developmentally regulated protein is
RT homologous to eukaryotic histone H1."
RL Proc. Natl. Acad. Sci. U.S.A. 88:3937-3941(1991).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=D/W-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: MIGHT HAVE A ROLE ANALOGOUS TO THAT OF EUKARYOTIC
CC HISTONE PROTEINS.
CC -1- DEVELOPMENTAL STAGE: SPECIFIC TO THE EB (ELEMENTARY BODY) FORM
CC IN THE LIFE CYCLE OF CHLAMYDIAE.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY. HCT SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57311; CAA40563.1; -.
DR EMBL: M60902; AAA23129.1; -.
DR EMBL: AE001345; AAC68338.1; -.
DR PIR: S16152; S16152.
DR PIR: A39396; A39396.
KW DNA-binding; Repeat; Complete proteome.
FT VARIANT 45 I -> T (IN STRAIN L2/434/BU).
FT VARIANT 83 A -> T (IN STRAIN L2/434/BU).
SQ SEQUENCE 125 AA; 13681 MW; F42AC6CE5BA822D3 CRC64;

Query Match 88.9%; Score 24; DB 1; Length 125;
Best Local Similarity 83.3%; Pred. No. 58;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
Db 69 AKKARA 74

RESULT 8
H2A_TRYCR
ID H2A_TRYCR STANDARD: PRT; 135 AA.
AC P35066;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Histone H2A.
OS Trypanosoma cruzi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxId=5693;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94359525; PubMed=8078513;
RA Puerta C., Martin J., Alonso C., Lopez M.C.;
RT "Isolation and characterization of the gene encoding histone H2A from
RT Trypanosoma cruzi."
RL Mol. Biochem. Parasitol. 64:1-10(1994).
RN [2]
RP REVISIONS TO 5; 92 AND 106.
RA Lopez Lopez M.C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H2A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X67287, CAA47703.2; -
 DR PIR: S25119; S25119.
 DR PIR: S33211; S33211.
 DR InterPro: IPR002119; Histone_H2A.
 DR InterPro: IPR004822; Histone_core.
 DR Pfam: PF00125; histone.1.
 DR PRINTS: PR00620; HISTONEH2A.
 DR PRODOM: PD000522; HISTONE_H2A; 1.
 DR SMART: SM00414; H2A; 1.
 DR PROSITE: PS00046; HISTONE_H2A; 1.
 DR Chromosomal protein; Nucleosome core; DNA-binding.
 SO SEQUENCE 135 AA; 1436 MW; DC41BCE5983EA93B CRC64;

Query Match
 Best Local Similarity 88.9%; Score 24; DB 1; Length 135;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKARA 6
 Db 126 SKKARA 131

RESULT 9
 CLSP_HUMAN STANDARD; PRT; 146 AA.
 AC Q9NZT1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Calmodulin-like skin protein.
 GN CLSP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Skin;
 RA MEDLINE=2023936; PubMed=10777582;
 RA Meul B., Bernard D., Simonetti L., Bernard M.A., Schmidt R.;
 RT "Identification and cloning of a new calmodulin-like protein from
 RT human epidermis.";
 RL J. Biol. Chem. 275:12841-12847(2000).
 CC -1- FUNCTION: BINDS CALCIUM. MAY BE INVOLVED IN TERMINAL
 CC DIFFERENTIATION OF KERATINOCYTES.
 CC -1- SUBUNIT: ASSOCIATES WITH TRANSGLUTAMINASE 3.
 CC -1- TISSUE SPECIFICITY: PARTICULARLY ABUNDANT IN THE EPIDERMIS WHERE
 CC ITS EXPRESSION IS DIRECTLY RELATED TO KERATINOCYTE
 CC DIFFERENTIATION. VERY LOW EXPRESSION IN LUNG.
 CC -1- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF172852; AAF66821.1; -
 DR HSSP: P02593; 1FW4.
 DR MIM: 605183; -
 DR InterPro: IPR002048; EF-hand.
 DR Pfam: PF00036; efhand; 4.
 DR PRODOM: PD000012; EF-hand; 2.
 DR SMART: SM00054; EFh; 4.
 DR PROSITE: PS00018; EF_HAND; 4.
 DR Calcium-binding; Repeat.
 KW CA_BIND 21 EF-HAND 1 (POTENTIAL).
 RP

FT CA_BIND 57 68 EF-HAND 2 (POTENTIAL).
 FT CA_BIND 91 102 EF-HAND 3 (POTENTIAL).
 FT CA_BIND 127 138 EF-HAND 4 (POTENTIAL).
 SQ SEQUENCE 146 AA; 15920 MW; 6F565F8E04B94CC CRC64;

Query Match
 Best Local Similarity 88.9%; Score 24; DB 1; Length 146;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKARA 6
 Db 73 AKKARA 78

RESULT 10
 HI_ECHCR STANDARD; PRT; 171 AA.
 ID HI_ECHCR
 AC P02257;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Histone H1, gonadal (Fragment).
 OS Echinolampas crassa (Sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Atelostomata; Cassiduloida;
 OC Echinolampadidae; Echinolampas.
 OX NCBI_TaxID=7646;
 RN [1]
 RP SEQUENCE.
 RA Strickland W.N., Strickland M., von Holt C.;
 RT "A comparison of the amino acid sequences of histones H1 from the
 RT sperm of Echinolampas crassa and Parichnus angulosus.";
 RL Biochim. Biophys. Acta 700:127-129(1982).
 CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
 CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: SPERM.
 CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
 CC PIR: A02587; HSUR1E.
 DR HSSP: P02259; 1HST.
 DR InterPro: IPR001386; Histone_H1/H5.
 DR Pfam: PF00538; linker_histone; 1.
 DR PRODOM: PD000373; linkerhist_N; 1.
 DR SMART: SM00526; H15; 1.
 KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
 KW Sperm.
 FT NON_TER 171 171
 SO SEQUENCE 171 AA; 18309 MW; 34DE1B20013E06BB CRC64;

Query Match
 Best Local Similarity 88.9%; Score 24; DB 1; Length 171;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKKARA 6
 Db 135 AKKARA 140

RESULT 11
 YLX9_CAEEL STANDARD; PRT; 176 AA.
 ID YLX9_CAEEL
 AC P46505;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Hypothetical 20.5 kDa protein F23P12.9 in chromosome III.
 GN F23P12.9.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-Bristol NZ;
 RA Du 2.1;
 RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: SOME, TO THE C-TERMINAL OF RAT DBP AND TER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U12965; AAA20611.1; -
 DR HSSP: P03069; 2DGC.
 DR WormPep: P23F12.9; CE01256.
 DR InterPro: IPR004827; TF_bZIP.
 DR SMART: SM00338; BRLZ; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 176 AA; 20468 MW; 96C763D2618E0434 CRC64;
 Query Match 88.9%; Score 24; DB 1; Length 176;
 Best Local Similarity 83.3%; Pred. No. 79;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AKKARA 6
 DB 79 AKKSRA 84
 RESULT 12
 ID R4_CYAME STANDARD; PRT; 191 AA.
 AC 022020;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Chloroplast 30S ribosomal protein S4.
 GN RPS4.
 OS Cyanidioschyzon merolae.
 OC Chloroplast.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Cyanidioschyzon.
 OX NCBI_TaxID=45157;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ohta N.;
 RT "Analysis of a plastid gene cluster reveals a close relationship
 RT between Cyanidioschyzon and Cyanidium."
 RL J. Plant Res. 110:235-245(1997).
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 1 S4 RNA-BINDING DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D63675; BAA22816.1; -
 DR HSSP: P81288; 1C05.
 DR InterPro: IPR001912; Ribosomal_S4.
 DR InterPro: IPR002942; S4.
 DR Pfam: PF00163; Ribosomal_S4; 1.
 DR Pfam: PF01479; S4; 1.
 DR SMART: SM00363; S4; 1.
 DR TIGRfams: TIGR01017; rpsd_bact; 1.
 DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
 KW Ribosomal protein; rRNA-binding; Chloroplast.
 FT DOMAIN 92 139 RNA-BINDING (S4 TYPE).

SQ SEQUENCE 191 AA; 21758 MW; A0CDF4B815DF5E61 CRC64;
 Query Match 88.9%; Score 24; DB 1; Length 191;
 Best Local Similarity 83.3%; Pred. No. 85;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AKKARA 6
 DB 75 AKKSRA 80
 RESULT 13
 ID Y981_METJA STANDARD; PRT; 204 AA.
 AC 058390;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein M00981.
 GN M00981.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=9633799; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Ullrich T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Mese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii."
 RL Science 273:1058-1073(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U67541; AAB98984.1; -
 DR TIGR: M00981; -
 KW Hypothetical protein; Complete proteome.
 FT DOMAIN 23 29 LYS-RICH.
 FT DOMAIN 163 181 LYS-RICH.
 SQ SEQUENCE 204 AA; 23595 MW; 6AB7CA2A26421FB6 CRC64;
 Query Match 88.9%; Score 24; DB 1; Length 204;
 Best Local Similarity 83.3%; Pred. No. 91;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AKKARA 6
 DB 172 AKKARA 177
 RESULT 14
 ID YPDC_BACSU STANDARD; PRT; 218 AA.
 AC P50738;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ypdC.
 FT ypdC.

OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / Matburg;
 RX MEDLINE=96349105; PubMed=8760912;
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
 RA Seror P.;
 RT "Sequence analysis of the Bacillus subtilis chromosome region between
 the serA and kds loci cloned in a yeast artificial chromosome.";
 RL Microbiology 142:2005-2016(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Berrtero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devant K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaere-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Kocha E., Koebe B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Sciffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis".
 RL Nature 390:249-256(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L47648; AAC83956.1; -
 DR EMBL: Z99115; CAB14210.1; -
 DR EMBL: Z99116; CAB14226.1; -
 DR Subtilisin; Bg11438; ypdC.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 30 50 POTENTIAL.
 FT TRANSMEM 65 85 POTENTIAL.
 FT TRANSMEM 127 147 POTENTIAL.
 FT TRANSMEM 180 200 POTENTIAL.
 SQ SEQUENCE 218 AA; 24719 MW; 88556D50863E14BC CRC64;

Query Match 88.9%; Score 24; DB 1; Length 218;
 Best Local Similarity 83.3%; Pred. No. 96;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AKKARA 6

Db 207 AKKARS 212
 |||||:
 RESULT 15
 H47_STELP
 ID H47_STELP STANDARD; PRY; 235 AA.
 AC Q41348;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Hypothetical protein H47 (Fragment).
 OS Stellaria longipes.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Caryophyllidae; Caryophyllales; Caryophyllaceae; Stellaria.
 OX NCBI_TaxID=19744;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RA Zhang X.H.;
 RL Submitted (MAY-1993) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0019 (SNZ) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X71601; CAA50602.1; ALT. INIT.
 DR InterPro: IPR003009; FEM enzyme.
 DR InterPro: IPR001852; SnzIP/SnZ.
 DR Pfam: PF01680; SNZ; 1.
 DR ProDom: PD004958; SnzIP/SnZ; 1.
 DR PROSITE: PS01235; UPF0019; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 235 AA; 25450 MW; 7009E5A53E2D0E0C CRC64;

Query Match 88.9%; Score 24; DB 1; Length 235;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AKKARA 6
 |||||:
 Db 184 AKKARA 189

Search completed: April 8, 2003, 11:24:36
 Job time : 7.65517 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:12:02 ; Search time 4.65517 Seconds

(without alignments)
53.458 Million cell updates/sec

Title: US-09-496-391-16

Perfect score: 27

Sequence: 1 ARAKKA 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	27	100.0	161 1 SECE_MYCTU	P96929 mycobacteri
2	27	100.0	320 1 MEC3_CAEVU	P34765 caenorhabdi
3	27	100.0	410 1 CGEI_HUMAN	P24864 homo sapien
4	27	100.0	424 1 BCHN_RHOCA	P26164 rhodobacter
5	24	88.9	97 1 Y13A_BPT4	P17310 bacterioph
6	24	88.9	171 1 H1_ECHCR	P02257 echinolam
7	24	88.9	180 1 RL17_MYCTU	P06323 mycobacteri
8	24	88.9	206 1 H1_ONCMY	P06350 oncorhynch
9	24	88.9	206 1 R131_BRANA	P41128 brassica na
10	24	88.9	206 1 R132_BRANA	P41129 brassica na
11	24	88.9	206 1 R133_ARATH	P41127 arabidopsi
12	24	88.9	209 1 H1A_XENLA	P06892 xenopus lae
13	24	88.9	217 1 H1_ANAPL	P05426 anas platyr
14	24	88.9	217 1 H1_CHICK	P05987 gallus gall
15	24	88.9	218 1 H1R_CHICK	P08288 gallus gall
16	24	88.9	219 1 H110_CHICK	P08286 gallus gall
17	24	88.9	222 1 H15_MOUSE	P43276 mus muscul
18	24	88.9	223 1 H103_CHICK	P08285 gallus gall
19	24	88.9	224 1 H1LL_CHICK	P08287 gallus gall
20	24	88.9	225 1 H15_HUMAN	P16401 homo sapien
21	24	88.9	245 1 H1_MAIZE	P23444 zea mays (m
22	24	88.9	248 1 H1_PANAN	P02255 parecchius
23	24	88.9	255 1 TPIS_YERPE	P08169 yerushla pe
24	24	88.9	297 1 CEBB_RAT	P21272 rattus norv
25	24	88.9	344 1 YK61_YEAST	P3160 saccharomyc
26	24	88.9	353 1 MRAY_HELPJ	P09191 heliobacte
27	24	88.9	353 1 MRAY_HELPJ	P09191 heliobacte
28	24	88.9	366 1 T2AA_DROME	P52654 drosophila
29	24	88.9	376 1 G4P1_YEAST	P46672 saccharomyc
30	24	88.9	399 1 OYEL_YEAST	P00899 saccharomyc
31	24	88.9	412 1 AFGM_PYROO	P05742 pyrococcus
32	24	88.9	416 1 PROA_LEPXA	P94872 leptospira
33	24	88.9	417 1 FRZC_MYXXA	P43500 myxococcus

34	24	88.9	425 1 PRS7_PROBE	O64982 prunus pers
35	24	88.9	426 1 PRS7_SPIOL	O41365 spinachia ol
36	24	88.9	428 1 BCHN_RHOCA	O9-ftd4 rhodobacter
37	24	88.9	434 1 PEL_TILLO	P40973 lilium long
38	24	88.9	452 1 PCNB_HAETN	P44439 haemophilus
39	24	88.9	505 1 MALO_STRPN	P29851 streptococc
40	24	88.9	525 1 SPL_RARFA	P05308 rarebacter
41	24	88.9	602 1 DCPL_ORYSA	P51847 oryza sativ
42	24	88.9	618 1 ORC2_DROME	O24168 drosophila
43	24	88.9	621 1 HEM1_AGABT	O92403 agaricus bi
44	24	88.9	639 1 GYRB_HALSO	P21558 halofetax s
45	24	88.9	659 1 AMIA_STRPN	P18791 streptococc

ALIGNMENTS

RESULT 1
SECE_MYCTU
ID SECE_MYCTU STANDARD: PRT; 161 AA.
AC P96929;
DF 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable preprotein translocase sece subunit.
GN SECE OR RV0638 OR MT0667 OR MTCY20H10.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RT Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
RL -1- FUNCTION: ESSENTIAL FOR PROTEIN EXPORT.
CC -1- SUBCELLULAR LOCATION: Tail-anchored membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SECE/SEC61-GAMMA FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: 292772; CAB07097.1; -;
CC EMBL: AE006961; -; NCOT_ANNOTATED_CDS.
CC EMBL: AE006962; AAK44892.1; ALT_INIT.
CC TIGR: MT0667; -;
CC Tuberculin; RV0638; -;

```

DR InterPro: IPR001901; SecE.
DR InterPro: IPR004819; SecE_bac.
DR Pfam: PF00584; SecE; 1.
DR ProDom: PD005139; SecE_bac; 1.
DR TIGRFAMs: TIGR00964; Jao051606; 1.
DR PROSITE: PS01067; SECE_SEC61G; 1.
KW Protein transport; Translocation; Transmembrane; Complete proteome.
FT TRANSMEM 133 153 POTENTIAL.
SQ SEQUENCE 161 AA; 16956 MW; 9AC5CE0696B9784E CRC64;

Query Match
Best Local Similarity 100.0%; Score 27; DB 1; Length 161;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 86 ARAKKA 91

RESULT 2
MEC3_CAEVU STANDARD; PRT; 320 AA.
AC P34765;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mechanosensory protein 3.
GN MEC-3.
OS Caenorhabditis vulgatis.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=31233;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92084094; PubMed=1684166;
RA Way J.C., Wang L., Run J.O., Wang A.;
RT "The mec-3 gene contains cis-acting elements mediating positive and
RT negative regulation in cells produced by asymmetric cell division in
RT Caenorhabditis elegans."
RL Genes Dev. 5:2199-2211(1991).
CC -1- FUNCTION: MEC-3 SPECIFIES DIFFERENTIATION OF THE SET OF SIX TOUCH
CC RECEPTOR NEURONS. BINDS COOPERATIVELY AS A HETERODIMER WITH UNC-86
CC TO SITES IN THE MEC-3 GENE PROMOTER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63956; CAA45377.1; -
DR PIR: A39479; A39479.
DR HSSP: P06601; 1FJL.
DR TRANSFAC: T01970; -
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR001781; LIM.
DR Pfam: PF00046; homeobox; 1.
DR ProDom: PD000010; Homeobox; 1.
DR ProDom: PD000094; LIM; 2.
DR SMART: SM00389; HOX; 1.
DR SMART: SM00132; LIM; 2.
DR PROSITE: PS00027; HOMEBOX_1; FALSE_NEG.
DR PROSITE: PS00071; HOMEBOX_2; 1.
DR PROSITE: PS00478; LIM_DOMAIN_1; 2.
DR PROSITE: PS00023; LIM_DOMAIN_2; 2.
KW Homeobox; DNA-binding; Nuclear protein; Developmental protein;

```

```

KW Repeat; LIM domain; Metal-binding; Zinc.
FT .DOMAIN 29 79 LIM 1.
FT .DOMAIN 89 145 LIM 2.
FT DNA_BIND 216 275 HOMEBOX.
FT DOMAIN 314 320 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 320 AA; 36849 MW; AF98BD31C282F8B6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 27; DB 1; Length 320;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 245 ARAKKA 250

RESULT 3
CGEL_HUMAN STANDARD; PRT; 410 AA.
AC P24864; Q14091; Q92501;
DT 01-MAR-1992 (Rel. 21, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE G1/S-specific cyclin E1.
GN CCNE1 OR CCNE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 8-410 FROM N.A.
RA MEDLINE=92005673; PubMed=1833068;
RA Koff A., Cross F., Fisher A., Schumacher J., Le Guellec K.,
RA Philippe M., Roberts J.M.;
RT "Human cyclin E, a new cyclin that interacts with two members of the
RT CDC2 gene family."
RL Cell 66:1217-1228(1991).
RN [2]
RP SEQUENCE OF 8-410 FROM N.A.
RA MEDLINE=92005671; PubMed=1833066;
RA Lew D.J., Dulic V., Reed S.I.;
RT "Isolation of three novel human cyclins by rescue of G1 cyclin (Cln)
RT function in yeast."
RL Cell 66:1197-1206(1991).
RN [3]
RP SEQUENCE OF 1-42 FROM N.A.
RA MEDLINE=9626347; PubMed=8649818;
RA Geng Y., Eaton E.N., Picon M., Roberts J.M., Lundberg A.S.,
RA Gifford A., Sardet C., Weinberg R.A.;
RT "Regulation of cyclin E transcription by E2Fs and retinoblastoma
RT protein."
RL Oncogene 12:1173-1180(1996).
RN [4]
RP SEQUENCE OF 281-370 FROM N.A.
RA Li H., Labat J.M., Valentine M., Saito M., Reed S.I., Look T.,
RA Kidd V.J.;
RT Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP ALTERNATIVE SPLICING.
RA MEDLINE=94266983; PubMed=8207080;
RA Sewing A., Roenicke V., Buerger C., Funk M., Mueller R.;
RT "Alternative splicing of human cyclin E."
RL J. Cell Sci. 107:581-588(1994).
RN [6]
RP PHOSPHORYLATION OF THR-395.
RA MEDLINE=97015119; PubMed=8861947;
RA Won K.A., Reed S.I.;
RT "Activation of cyclin E/CDK2 is coupled to site-specific
RT autophosphorylation and ubiquitin-dependent degradation of cyclin E."
RL EMBO J. 15:4182-4193(1996).
RN [7]
RP TISSUE SPECIFICITY.
RA MEDLINE=99054678; PubMed=9840943;

```

RA Zariwala M., Liu J., Xiong Y.;
RT "Cyclin E2, a novel human G1 cyclin and activating partner of CDK2 and
RT CDK3, is induced by viral oncoproteins.";
RL Oncogene 17:2787-2798(1998).
CC
CC -1- FUNCTION: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G1/S
CC (START) TRANSITION.
CC -1- SUBUNIT: INTERACTS WITH A MEMBER OF THE CDK2/CDK PROTEIN KINASES
CC TO FORM A SERINE/THREONINE KINASE HOLOENZYME COMPLEX. THE CYCLIN
CC SUBUNIT IMPARTS SUBSTRATE SPECIFICITY TO THE COMPLEX. INTERACTS
CC WITH RETINOBLASTOMA BINDING PROTEIN 3 AND RETINOBLASTOMA-LIKE
CC PROTEIN 1.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: E1L (SHOWN HERE) AND E1S; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. E1S LACKS 49 RESIDUES WITHIN THE
CC CYCLIN BOX AND CANNOT COMPLEX WITH CDK2.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS AND PLACENTA. LOW
CC LEVELS IN BRONCHIAL EPITHELIAL CELLS.
CC -1- PTM: PHOSPHORYLATION BY CDK2 TRIGGERS ITS RELEASE FROM CDK2 AND
CC DEGRADATION VIA THE UBIQUITIN PROTEASOME PATHWAY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN E SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: M73812; -; NOT_ANNOTATED_CDS.
DR EMBL: M74093; -; NOT_ANNOTATED_CDS.
DR EMBL: X95406; CA64687.1; -.
DR EMBL: X95406; CA64688.1; -.
DR EMBL: U40788; AAA83269.1; -.
DR EMBL: U40787; AAA83269.1; JOINED.
DR PIR: A40270; A40270.
DR PIR: C40268; C40268.
DR Genew: HGNC:1589; CCNEL.
DR MIM: 123837; -.
DR InterPro: IPR004366; Cyclin.
DR InterPro: IPR004367; Cyclin_Cterm.
DR Pfam: PF00134; cyclin; 1.
DR Pfam: PF02984; cyclin_C; 1.
DR SMART: SM00385; CYCLIN; 1.
DR PROSITE: PS00292; CYCLINS; 1.
KW Cyclin: cell cycle; cell division; Alternative splicing;
KW Phosphorylation; Nuclear protein.
FT MOD_RES 395 395 PHOSPHORYLATION.
FT VARSPIC 154 196 MISSING (IN ISOFORM E1S).
FT CONFLICT 281 281 L -> M (IN REF. 4).
SQ SEQUENCE 410 AA; 47077 MW; 424DF0B253B7047E CRC64;

Query Match 100.0%; Score 27; DB 1; Length 410;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 373 ARAKKA 378

RESULT 4
BCNN_RHOCA
ID BCNN_RHOCA STANDARD; PRT; 424 AA.
AC P26164;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Light-independent protochlorophyllide reductase subunit N
DE (Ec 1.18.-.-) (LI-FOR subunit N) (DPOR subunit N).
GN BCNN.
OS Rhodospirillum rubrum (Rhodospirillum rubrum).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;

CC Rhodospirillum rubrum.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003 / St Louis;
RX MEDLINE=93224465; PubMed=8385667;
RA Burke D.H., Alberti M., Hearst J.E.;
RT "bchN/bchH bacteriochlorophyll synthase genes of Rhodospirillum rubrum
RT and identification of the third subunit of light-independent
RT protochlorophyllide reductase in bacteria and plants.";
RT J. Bacteriol. 175:2414-2422(1993).
RL [2]
RN CHARACTERIZATION.
RP STRAIN=SB1003 / CB1029;
RC MEDLINE=20378986; PubMed=10811655;
RX Fujita Y., Bauer C.E.;
RA Fujita Y., Bauer C.E.;
RT "Reconstitution of light-independent protochlorophyllide reductase
RT from purified bchL and bchN-bchH subunits. In vitro confirmation of
RT nitrogenase-like features of a bacteriochlorophyll biosynthesis
RT enzyme.";
RL J. Biol. Chem. 275:23583-23588(2000).
RN [3]
RP CHARACTERIZATION.
RA Fujita Y.;
RL Unpublished observations (JUL-2001).
CC -1- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
CC protochlorophyllide (pchlide) to form chlorophyllide a (Chlide).
CC The bchN-bchH pair binds pchlide. This reaction is light-
CC independent.
CC -1- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.
CC -1- SUBUNIT: Protochlorophyllide reductase is thought to be composed
CC of three subunits: bchL, bchN and bchH. Could form a
CC heterotrimer of two bchH and two bchN subunits.
CC -1- SIMILARITY: BELONGS TO THE BCNN / CHLN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: Z11165; CA77526.1; -.
DR PIR: S17810; S17810.
DR PIR: B49851; B49851.
DR TIGRFAAS: TIGR01279; DPOR_bchN; 1.
KW Photosynthesis; Bacteriochlorophyll biosynthesis; Oxidoreductase.
SQ SEQUENCE 424 AA; 45829 MW; C90C75233802834D CRC64;

Query Match 100.0%; Score 27; DB 1; Length 424;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 279 ARAKKA 284

RESULT 5
Y13A_BPT4
ID Y13A_BPT4 STANDARD; PRT; 97 AA.
AC P17310; O9T0T5;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 11.1 kDa protein in Gp30-rIII intergenic region (ORF D).
GN Y13A OR 30.9 OR 31.-2.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10665;
RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=D;
RX MEDLINE=90301484; PubMed=2362813;
RA Prilipov A.G., Mesyanzhinov V.V., Aebi U., Kellenberger E.;
RT "Cloning and sequencing of bacteriophage T4 genes between map
RT positions 128.3-130.3";
RL Nucleic Acids Res. 18:3635-3635(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90332452; PubMed=2377483;
RA Raudonkiene A., Nivinskas R.;
RT "Nucleotide sequence of bacteriophage T4 gene 31 region.";
RL Nucleic Acids Res. 18:4280-4280(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92267389; PubMed=1587487;
RA Raudonkiene A., Nivinskas R.;
RT "Gene rIII is the nearest downstream neighbour of bacteriophage T4
RT gene 31.";
RL gene 114:85-90(1992).
RN [4]
RP SEQUENCE FROM N.A.
RA Kuter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Sliham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -i CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-26 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X17657; CAA35653.1; -
DR EMBL: M37882; AAA32508.1; ALT_INIT.
DR EMBL: X54536; CAA38407.1; ALT_INIT.
DR EMBL: AF158101; AAD42450.1; ALT_INIT.
DR PIR: J00527; J00527.
DR PIR: S26172; S26172.
KW Hypothetical protein.
SQ SEQUENCE 97 AA; 11087 MW; 50C87172C3C22CA51 CRC64;

Query Match      88.9%; Score 24; DB 1; Length 97;
Best Local Similarity 83.3%; Pred. No. 45;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
   1:||||
Db 44 ARAKKA 49

RESULT 6
HI_ECHCR STANDARD: PRT; 171 AA.
AC P02257;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1, gonadal (Fragment).
OS Echinolampas crassa (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Echinoidea; Atelostomata; Cassiduloida;
OC Echinolampidae; Echinolampas.
NX NCBI-Taxid=7646;
RN [1]
RP SEQUENCE.
RA Strickland W.N., Strickland M., von Holt C.;
RT "A comparison of the amino acid sequences of histones H1 from the
RT sperm of Echinolampas crassa and Parachinus angulosus.";
RL Biochim. Biophys. Acta 700:127-129(1982).
```

```

CC -i FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -i SUBCELLULAR LOCATION: Nuclear.
CC -i TISSUE SPECIFICITY: SPERM.
CC -i SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR: A02587; HSURE.
DR HSP: P02259; HST.
DR InterPro: IPR001386; Histone_H1/H5.
DR Pfam: PF00538; linker_histone_1.
DR ProDom: PD000373; Linkerhist_N; 1.
DR SMART: SM00526; H1S; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Sperm.
FT NON_TER
SQ SEQUENCE 171 AA; 18309 MW; 34DE1B20013E06BB CRC64;

Query Match      88.9%; Score 24; DB 1; Length 171;
Best Local Similarity 83.3%; Pred. No. 75;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
   1:||||
Db 133 ARAKKA 138

RESULT 7
RL17_MYCTU STANDARD: PRT; 180 AA.
ID 006323;
AC 006323;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L17.
GN RPLQ OR RVJ3456C OR MT3563 OR MTCY13E12.09C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI-Taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagsis K., Krogh A., McLaren J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Petersmann R.D., Allard D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -i SIMILARITY: BELONGS TO THE L17P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```



```

CC -----
DR EMBL; Z95390; CAB08722.1; -.
DR EMBL; AE007160; AAK47902.1; -.
DR TIGR; MT3563; -.
DR TubercuList; RV3456c; -.
DR InterPro; IPR000456; Ribosomal_L17.
DR Pfam; PF01196; Ribosomal_L17; 1.
DR ProDom; PD004277; Ribosomal_L17; 1.
DR TIGRFAMs; TIGR00059; L17; 1.
DR PROSITE; PS01167; RIBOSOMAL_L17; 1.
KW Ribosomal protein; Complete proteome.
FT CONFLICT 175 175 A->T (IN REF. 2).
SQ SEQUENCE 180 AA; 19475 MW; ACA33FB5ED6DE99F CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 180;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 136 ARAKKA 141

RESULT 8
H1_ONCMY STANDARD; PRT; 206 AA.
ID H1_ONCMY
AC P06350;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
NC NCB1_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85264847; PubMed=6443128;
RA Mesquita J., Connor W., Winkfein R.J., Dixon G.H.;
RL "An H1 histone gene from rainbow trout (Salmo gairdneri).";
J. Mol. Evol. 21:209-219(1985).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02624; CAB37646.1; -.
DR PIR; A02584; HSTR1R.
DR HSSP; P08287; IGHC.
DR InterPro; IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone; 1.
DR ProDom; PD000373; Linker_histone; 1.
DR SMART; SM00526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation.
FT INT_MET 0 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT DOMAIN 27 100 GLOBULAR.
SQ SEQUENCE 206 AA; 20672 MW; 72C440798066716C CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 206;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ARAKKA 6
Db 196 ARAKKA 201

RESULT 9
R131_BRANA STANDARD; PRT; 206 AA.
ID R131_BRANA
AC P41128;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 60S ribosomal protein L13-1 (Cold induced protein C24A).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
NC NCB1_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Samourai;
RX MEDLINE=94122385; PubMed=8292785;
RA Saez-Vasquez J., Raynal M., Meza-Basso L., Delseny M.;
RL "Two related, low-temperature-induced genes from Brassica napus are
RL homologous to the human tumour bcl1 (breast basic conserved) gene.";
Plant Mol. Biol. 23:1211-1221(1993).
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22618; CAA80341.1; -.
DR PIR; S37132; S37132.
DR InterPro; IPR001380; Ribosomal_L13E.
DR Pfam; PF01294; Ribosomal_L13e; 1.
DR ProDom; PD004443; Ribosomal_L13e; 1.
DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein; Multigene family.
SQ SEQUENCE 206 AA; 23633 MW; DE7CCD26919431AF CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 206;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 192 ARAKKA 197

RESULT 10
R132_BRANA STANDARD; PRT; 206 AA.
ID R132_BRANA
AC P41129;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 60S ribosomal protein L13-2 (Cold induced protein C24B).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
NC NCB1_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Samourai;
RX MEDLINE=94122385; PubMed=8292785;
RA Saez-Vasquez J., Raynal M., Meza-Basso L., Delseny M.;
RL "Two related, low-temperature-induced genes from Brassica napus are

```

```

RT homologous to the human tumour bcl1 (breast basic conserved) gene."
RL Plant Mol. Biol. 23:1211-1221(1993).
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z22620; CAA80343.1; -
DR EMBL: Z22619; CAA80342.1; -
DR PIR: S37134; S37134.
DR InterPro: IPR001380; Ribosomal_L13E.
DR Pfam: PF01294; Ribosomal_L13e; 1.
DR ProDom: PD004443; Ribosomal_L13E; 1.
DR PROSITE: PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein; Multigene family.
SQ SEQUENCE 206 AA; 23656 MW; C6929653484851E6 CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 206;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
DB 192 ARAKKA 197

RESULT 11
RL3-ARATH STANDARD; PRT; 206 AA.
ID RL3-ARATH
AC P41127;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L13 (BBL1 protein homolog).
GN RPL13 OR BBL1 OR AT3G49010 OR T2J13.150.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
ON [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=94215905; PubMed=8163191;
RA Bertuche N., Leung J., Giraudat J.;
RT "Conservation of the human breast basic conserved 1 gene in the plant
RT kingdom: characterization of a cDNA clone from Arabidopsis
RT thaliana."
RL Gene 141:211-214(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unseld M.,
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Delesny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choise N., Artiguenave F., Robert C., Brottier P.,
RA Winkler P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurnbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Smidoni B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T., H., Nordstiek G.,
RA Reichelt J., Schaefer M., Schoen O., Barques M., Terol J., Clement J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Landie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Collet A., Casacuberta E.,
RA Montfort A., Argitau A., Flores M., Liguori R., Vitale D.,

```

```

RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Uterback T., Fujii C.Y., Shea T.P.,
RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltcher J., Sellers P., Gill J.E., Feldlyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
RA Matsubae A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana."
RL Nature 408:820-822(2000).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X75162; CAA53005.1; -
DR EMBL: AL132967; CAB62009.1; -
DR PIR: S37271; S37271.
DR InterPro: IPR001380; Ribosomal_L13E.
DR Pfam: PF01294; Ribosomal_L13e; 1.
DR ProDom: PD004443; Ribosomal_L13E; 1.
DR PROSITE: PS01104; RIBOSOMAL_L13E; 1.
KW Ribosomal protein.
SQ SEQUENCE 206 AA; 23767 MW; C67BA5E534114C2 CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 206;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
DB 192 ARAKKA 197

RESULT 12
HLA_XENLA STANDARD; PRT; 209 AA.
ID HLA_XENLA
AC P06892;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1A.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
ON [1]
RP SEQUENCE FROM N.A. (GENE CLUSTER X1H3).
RX MEDLINE=86037224; PubMed=3863963;
RA Perry M., Thomson G.H., Roeder R.G.;
RT "Genomic organization and nucleotide sequence of two distinct histone
RT gene clusters from Xenopus laevis. Identification of novel conserved
RT upstream sequence elements."
RL J. Mol. Biol. 185:479-499(1985).
RN [2]
RP SEQUENCE FROM N.A.
RA Perry M., Thomson G.H., Roeder R.G.;
RT "Genomic organization and nucleotide sequence of two distinct histone
RT gene clusters from Xenopus laevis: Identification of novel conserved
RT upstream sequence elements."
RL J. Biol. Chem. 260:479-499(1985).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF

```

```
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X03018; CAZ26815.1; -
DR EMBL; M21287; AAA49767.1; -
DR PIR; F24510; HXX11A.
DR HSSP; P08287; IGHC.
DR InterPro: IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone; 1.
DR PRODOM; PD000373; linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
FT INIT_MET 0
FT DOMAIN 41 112 GLOBULAR.
SQ SEQUENCE 209 AA; 21242 MW; 8537E9272678DF1A CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 209;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ARAKKA 6
   1:||||
Db 199 AKAKKA 204

RESULT 13
H1_AKAPL STANDARD; PRT; 217 AA.
AC P09426;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.
OS Anas platyrhynchos (Domestic duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX NCBI_TaxID=8839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88036074; PubMed=2822942;
RA Tenjes R., Doecke D.;
RT "A highly conserved sequence in H1 histone genes as an
RT oligonucleotide hybridization probe: isolation and sequence of a duck
RT H1 gene";
RL J. Mol. Evol. 25:361-370(1987).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X06128; CAZ29495.1; -
DR PIR; S01262; S01262.
DR HSSP; P08287; IGHC.
DR InterPro: IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone; 1.
DR PRODOM; PD000373; linkerhist_N; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT DOMAIN 35 112 GLOBULAR.
SQ SEQUENCE 217 AA; 21750 MW; DED1926183438A29 CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 217;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ARAKKA 6
   1:||||
Db 207 AKAKKA 212

RESULT 14
H1_CHICK STANDARD; PRT; 217 AA.
AC P09987;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83238548; PubMed=6190814;
RA Sugarmen B.J., Dodgson J.B., Engel J.D.;
RT "Genomic organization, DNA sequence, and expression of chicken
RT embryonic histone genes.";
RL J. Biol. Chem. 258:9005-9016(1983).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; J00863; -; NOT_ANNOTATED_CDS.
DR PIR; A29179; HSC11.
DR HSSP; P08287; IGHC.
DR InterPro: IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone; 1.
DR PRODOM; PD000373; linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT DOMAIN 35 112 GLOBULAR.
SQ SEQUENCE 217 AA; 21750 MW; DED1926183438A29 CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 217;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ARAKKA 6
   1:||||
Db 207 AKAKKA 212
```

```
DR SMART; SM00526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT DOMAIN 35 112 GLOBULAR.
SQ SEQUENCE 217 AA; 21700 MW; F834CA74CA284951 CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 217;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ARAKKA 6
   1:||||
Db 207 AKAKKA 212

RESULT 14
H1_CHICK STANDARD; PRT; 217 AA.
AC P09987;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83238548; PubMed=6190814;
RA Sugarmen B.J., Dodgson J.B., Engel J.D.;
RT "Genomic organization, DNA sequence, and expression of chicken
RT embryonic histone genes.";
RL J. Biol. Chem. 258:9005-9016(1983).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; J00863; -; NOT_ANNOTATED_CDS.
DR PIR; A29179; HSC11.
DR HSSP; P08287; IGHC.
DR InterPro: IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone; 1.
DR PRODOM; PD000373; linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT DOMAIN 35 112 GLOBULAR.
SQ SEQUENCE 217 AA; 21750 MW; DED1926183438A29 CRC64;

Query Match
Best Local Similarity 88.9%; Score 24; DB 1; Length 217;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ARAKKA 6
   1:||||
Db 207 AKAKKA 212
```

```

RESULT 15
H1R.CHICK
ID H1R.CHICK STANDARD; PRT; 218 AA.
AC P08288;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.11R.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxId=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87250632; PubMed=3597432;
RA Coles L.S., Robins A.J., Madley L.K., Wells J.R.E.;
RT "Characterization of the chicken histone H1 gene complement.
  Generation of a complete set of vertebrate H1 protein sequences.";
RL J. Biol. Chem. 262:9656-9663(1987).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17020; AAA48790.1; -.
DR PIR; C28456; C28456.
DR HSSP; P08287; IGHC.
DR InterPro; IPR001386; Histone_H1/H5.
DR Pfam; PF00538; linker_histone_1.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H1S; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
FT INIT_MET 0
FT DOMAIN 37 110 GLOBULAR
SQ SEQUENCE 218 AA; 21672 MW; CB9724BFF14654A6 CRC64;

Query Match 88.9%; Score 24; DB 1; Length 218;
Best Local Similarity 83.3%; Pred. No. 94;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 208 ARAKKA 213

```

Search completed: April 8, 2003, 11:24:45
 Job time : 6.65517 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:17:38 ; Search time 21.3103 Seconds
(without alignments)
58.013 Million cell updates/sec

Title: US-09-496-391-16
Perfect score: 27
Sequence: 1 ARAKKA 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SPREMBL_21:*
- 2: sp.archaea:*
- 3: sp.bacteria:*
- 4: sp.fungi:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.potent:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.rviro:*
- 16: sp.bacteriap:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	217	16 Q8UDU9	Q8UDU9 agrobacteri
2	27	100.0	235	10 Q9FYT4	Q9FYT4 zea mays (m
3	27	100.0	359	2 Q85098	Q85098 pseudomonas
4	27	100.0	359	2 Q60239	Q60239 pseudomonas
5	27	100.0	375	11 Q8R5C3	Q8R5C3 mus musculus
6	27	100.0	415	16 Q914I2	Q914I2 pseudomonas
7	27	100.0	473	4 Q96FA5	Q96FA5 homo sapien
8	27	100.0	485	10 Q9S3H2	Q9S3H2 arabidopsis
9	27	100.0	498	17 Q8TP46	Q8TP46 methanosarc
10	27	100.0	715	10 Q9FYT6	Q9FYT6 zea mays (m
11	27	100.0	733	2 P77833	P77833 bordetella
12	27	100.0	753	16 Q9RDI9	Q9RDI9 streptomyce
13	27	100.0	880	5 Q961M1	Q961M1 drosophila
14	27	100.0	884	5 Q9VB71	Q9VB71 drosophila
15	27	100.0	987	4 Q9BU17	Q9BU17 homo sapien
16	27	100.0	998	4 Q9P211	Q9P211 homo sapien

17	27	100.0	1184	16 Q9F2P1	Q9F2P1 streptomyce
18	24	88.9	58	2 Q33937	Q33937 saccharopol
19	24	88.9	58	10 Q9ST54	Q9ST54 arabidopsis
20	24	88.9	59	9 Q8SBJ9	Q8SBJ9 bacterioph
21	24	88.9	65	16 Q8XQ9	Q8XQ9 ralsionia s
22	24	88.9	73	5 Q21411	Q21411 caenorhabdi
23	24	88.9	81	5 Q9XWY9	Q9XWY9 caenorhabdi
24	24	88.9	100	17 Q976E3	Q976E3 sulfolobus
25	24	88.9	118	5 Q96955	Q96955 plasmodium
26	24	88.9	120	2 Q9FBA3	Q9FBA3 bordetella
27	24	88.9	120	17 Q980V1	Q980V1 sulfolobus
28	24	88.9	137	16 Q98L34	Q98L34 rhizobium 1
29	24	88.9	140	5 Q9VWF8	Q9VWF8 rhizobium
30	24	88.9	149	2 Q8VLX5	Q8VLX5 xanthomonas
31	24	88.9	154	10 Q24227	Q24227 oryza sativ
32	24	88.9	160	16 Q9KQ32	Q9KQ32 vibrio chol
33	24	88.9	170	16 Q9F226	Q9F226 streptomyce
34	24	88.9	172	16 Q91093	Q91093 pseudomonas
35	24	88.9	191	5 Q6140	Q6140 mytilus edu
36	24	88.9	191	5 Q6141	Q6141 mytilus edu
37	24	88.9	191	5 Q6142	Q6142 mytilus edu
38	24	88.9	191	5 Q6143	Q6143 mytilus edu
39	24	88.9	191	5 Q6362	Q6362 mytilus edu
40	24	88.9	198	5 Q9VVR8	Q9VVR8 drosophila
41	24	88.9	200	5 Q969A1	Q969A1 haemaphysal
42	24	88.9	203	5 Q95SH2	Q95SH2 drosophila
43	24	88.9	206	10 Q9FF90	Q9FF90 arabidopsis
44	24	88.9	214	10 Q8RUS7	Q8RUS7 oryza sativ
45	24	88.9	217	16 Q82156	Q82156 yerstina pe

ALIGNMENTS

RESULT 1

ID	Q8UDU9	PRELIMINARY:	PRT:	217 AA.
AC	Q8UDU9:			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Transcriptional regulator, Teer family.			
GN	ATU2020 OR AGR_C_3662.			
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;			
OC	Rhizobiaceae; Rhizobium.			
OX	NCBI_TaxID=176299;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21608550; PubMed=11743193;			
RA	Wood D.W., Setubal J.C., Kaul R., Monks D.F., Kitajima J.P.,			
RA	Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,			
RA	Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D. St.,			
RA	Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,			
RA	Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,			
RA	Raymond C., Rouse G., Seemphimachak C., Wu Z., Romero P., Gordon D.,			
RA	Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,			
RA	Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,			
RA	Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,			
RA	Nester E.W.;			
RT	"The genome of the natural genetic engineer Agrobacterium tumefaciens			
RT	C58.";			
RL	Science 294:2317-2323(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21608551; PubMed=11743194;			
RA	Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,			
RA	Guoillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,			
RA	Homiel K., Gordon J., Vandin M., Iartchouk O., Epp A., Liu F.,			
RA	Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,			
RA	Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,			
RA	Cielo C., Slater S.;			
RT	"Genome sequence of the plant pathogen and biotechnology agent			

RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AE009153; AAL43012.1; -
DR EMBL; AE008118; AAK87774.1; -
KW Complete proteome.
SQ SEQUENCE 217 AA; 23795 MW; 8E828F2C350C949D CRC64;

Query Match 100.0%; Score 27; DB 16; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 190 ARAKKA 195

RESULT 2
ID 09FYT4 PRELIMINARY; PRT; 235 AA.
AC 09FYT4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CRS1 (Fragment).
GN CRS1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Till B.J., Schmitz-Linneweber C., Barkan A.;
RT "CRS1: A nuclear gene required for the splicing of the group II intron
in the chloroplast atpF gene."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF290416; AAG00597.1; -
FT NON_TER 235 235
SQ SEQUENCE 235 AA; 25751 MW; B516F563DA672575 CRC64;

Query Match 100.0%; Score 27; DB 10; Length 235;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 189 ARAKKA 194

RESULT 3
ID 085098 PRELIMINARY; PRT; 359 AA.
AC 085098;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Hrcv.
GN HRCV.
OS Pseudomonas syringae (pv. phaseolicola).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=319;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NPS3121;
RA Frederick R.D., Panopoulos N.J.;
RT "The hrpD locus of Pseudomonas syringae pv. phaseolicola encodes the
five most broadly conserved proteins in the signal peptide independent
RT (type III) protein export system common to plant and animal pathogenic
bacteria."
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043444; AAC25072.1; -
DR InterPro; IPR003439; ABC_Transporter.

DR InterPro; IPR002066; Bac_export_2.
DR Pfam; PF01312; Bac_export_2; 1.
DR PRINTS; PR00950; TYPE3IMSPROT.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
SQ SEQUENCE 359 AA; 39993 MW; 9CFD87BF2AEB56CF CRC64;

Query Match 100.0%; Score 27; DB 2; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 299 ARAKKA 304

RESULT 4
ID 060239 PRELIMINARY; PRT; 359 AA.
AC 060239;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HrpY.
GN HRPY.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=61;
RX MEDLINE=96025090; PubMed=7579617;
RA Huang H.C., Lin R.H., Chang C.J., Collmer A., Deng W.L.;
RT "The complete hrp gene cluster of Pseudomonas syringae pv. syringae 61
includes two blocks of genes required for harpinPs secretion that are
RT arranged colinearly with yersinia ysc homologs."
RL Mol. Plant Microbe Interact. 8:733-746(1995).
DR EMBL; U25812; AAB05078.1; -
DR InterPro; IPR002066; Bac_export_2.
DR Pfam; PF01312; Bac_export_2; 1.
DR PRINTS; PR00950; TYPE3IMSPROT.
SQ SEQUENCE 359 AA; 39984 MW; BE74AD3C0A8075E CRC64;

Query Match 100.0%; Score 27; DB 2; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 299 ARAKKA 304

RESULT 5
ID 08R5C3 PRELIMINARY; PRT; 375 AA.
AC 08R5C3;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 42.0 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC023044; AAH23044.1; -
KW Hypothetical protein.
SQ SEQUENCE 375 AA; 42013 MW; 07CD529833C46CBD CRC64;

Query Match 100.0%; Score 27; DB 11; Length 375;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 193 ARAKKA 198

RESULT 6

O914T2 PRELIMINARY; PRT; 415 AA.
AC O914T2
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ribonucleoside reductase, small chain.
GN NRDB OR PA1155.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Gabor R.L., Goltzy L., Tolentino E., Westbrock-Adamson S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Slier M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AF004545; AAC04544.1.
DR InterPro; IPR000358; RibonucL_redctse.
DR Pfam; PF00268; ribonucL_red_sm; 1.
KW Complete proteome.
SQ SEQUENCE 415 AA; 47386 MW; 803966F3DB2CCE6 CRC64;

Query Match 100.0%; Score 27; DB 16; Length 415;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 55 ARAKKA 60

RESULT 7

O96E45 PRELIMINARY; PRT; 473 AA.
AC O96E45
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Unknown (Protein for IMAGE:4475680) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RA Strausberg R.;
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012928; AA012928.1.
FT NON_TER
SQ SEQUENCE 473 AA; 52701 MW; 92D638B64153B00A CRC64;

Query Match 100.0%; Score 27; DB 4; Length 473;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 105 ARAKKA 110

RESULT 8

O9SJH2 PRELIMINARY; PRT; 485 AA.
AC O9SJH2
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative cytochrome P450.
GN ATG242850.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umayam U., Taiton L.J., Gill J.E.,
RA Adams M.D., Carreira A.J., Cressy T.H., Goodman H.M., Somerville C.R.,
RA Salzbarg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1. SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AC006931; AAD21724.1.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 485 AA; 55405 MW; 08B0B16474620F82 CRC64;

Query Match 100.0%; Score 27; DB 10; Length 485;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
Db 232 ARAKKA 237

RESULT 9

O8TP46 PRELIMINARY; PRT; 498 AA.
AC O8TP46
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein MA2073.
GN MA2073.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;

```

RX MEDLINE-21929760; PubMed-11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArillano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kueltner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL: AE010893; AAM05474.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 498 AA; 55934 MW; 2B1BB891ADB05AB8 CRC64;

Query Match 100.0%; Score 27; DB 17; Length 498;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 299 ARAKKA 304

RESULT 10
ID 09FYT6 PRELIMINARY; PRT; 715 AA.
AC 09FYT6;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE CRS1.
GN CRS1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Tili B.J., Schmitz-Lineweber C., Barkan A.;
RT "CRS1: A nuclear gene required for the splicing of the group II intron
RT in the chloroplast atpF gene.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF290414; AAG00595.1; -.
SQ SEQUENCE 715 AA; 81405 MW; 9ADC6ECE35F6C55F CRC64;

Query Match 100.0%; Score 27; DB 10; Length 715;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 189 ARAKKA 194

RESULT 11
ID P77833 PRELIMINARY; PRT; 733 AA.
AC P77833;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE RNA polymerase sigma factor.
GN RPOD.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=18323;
RX MEDLINE-97231336; PubMed-9076731;
RA Steffen P., Goyard S., Ullmann A.;
RT "The Bordetella pertussis sigma subunit of RNA polymerase confers
RT enhanced expression of the in Escherichia coli.";
RL Mol. Microbiol. 23:945-954(1997).
CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.
DR EMBL: U73858; AAC45085.1; -.
DR HSSP: P00579; 1SIG.
DR InterPro: IPR000943; Sigma_70.
DR Pfam: PF00140; sigma70; 1.
DR PRINTS: PR00046; SIGMA70FCT.
DR PROSITE: PS00715; SIGMA70_1; 1.
DR PROSITE: PS00716; SIGMA70_2; 1.
KW DNA-binding; DNA-directed RNA polymerase; Sigma factor;
KW Transcription regulation.
SQ SEQUENCE 733 AA; 81347 MW; B18730BC7461F703 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 733;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 47 ARAKKA 52

RESULT 12
ID 09RD19 PRELIMINARY; PRT; 753 AA.
AC 09RD19;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative deoxyribonuclease.
GN SC02737 OR SCC57A.08C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K.J., Harris D.;
RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RC STRAIN=A3(2);
RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RL [3]
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE-97000351; PubMed-8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M45;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,

```


RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RL coelicolor A3(2).";
 RU Nature 417:141-147(2002).
 DR EMBL; AL136519; CAB66276.1; -;
 DR InterPro; IPR003583; HHH.1.
 DR SMART; SM00278; Hhh1; 1.
 SQ SEQUENCE 753 AA; 81478 MW; A56EACAA23F9912A CRC64;

Query Match 100.0%; Score 27; DB 16; Length 753;
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
 DB 390 ARAKKA 395

RESULT 13

ID 0961M1 PRELIMINARY; PRT; 880 AA.
 AC 0961M1;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE GH16431P.
 GN CG6059.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY051511; AAK92935.1;
 DR FLYBase; FBgn0039491; CG6059.
 DR InterPro; IPR002007; Anim.peroxidase.
 SQ SEQUENCE 880 AA; 103622 MW; 9E13F28C6274EC6B CRC64;

Query Match 100.0%; Score 27; DB 5; Length 880;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
 DB 225 ARAKKA 230

RESULT 14

ID 09VB71 PRELIMINARY; PRT; 884 AA.
 AC 09VB71;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE CG6059 protein.
 GN CG6059.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;
 RA MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.C., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003759; AAF56672.1;
 DR FLYBase; FBgn0039491; CG6059.
 DR InterPro; IPR002007; Anim.peroxidase.
 SQ SEQUENCE 884 AA; 104078 MW; 929C17B8BC70F1BE CRC64;

Query Match 100.0%; Score 27; DB 5; Length 884;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
 DB 225 ARAKKA 230

RESULT 15

ID 09BU17 PRELIMINARY; PRT; 987 AA.
 AC 09BU17;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical 110.2 kDa protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC002575; AAH02575.1;
 DR InterPro; IPR001410; DEAD.

DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; helicase_C; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 987 AA; 110202 MW; 590EP94358750924 CRC64;

Query Match 100.0%; Score 27; DB 4; Length 987;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
DB 352 ARAKKA 357

Search completed: April 8, 2003, 11:28:26
Job time : 23.3103 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:11:45 ; Search time 25.3448 seconds
(without alignments)
31.545 Million cell updates/sec

Title: US-09-496-391-4
Perfect score: 27
Sequence: 1 AKKARA 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: A_Geneseq_101002.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	6	21	AA08139
2	27	100.0	12	21	AA08180
3	27	100.0	24	21	AA08181
4	27	100.0	30	21	AA08167
5	27	100.0	36	21	AA08169
6	27	100.0	89	22	AA08934
7	27	100.0	146	21	AA08754
8	27	100.0	156	23	AA08713
9	27	100.0	163	22	AA08757
10	27	100.0	163	22	AA04362

11	27	100.0	163	22	AA019929
12	27	100.0	537	18	AA034562
13	27	100.0	537	19	AA049866
14	27	100.0	1270	22	AA099141
15	27	100.0	1434	21	AA052028
16	27	100.0	1434	21	AA051657
17	27	100.0	1455	22	AA099140
18	27	100.0	1458	22	AA096825
19	24	88.9	8	23	AA074831
20	24	88.9	45	21	AA082572
21	24	88.9	66	21	AA082574
22	24	88.9	72	22	AA059512
23	24	88.9	92	22	AA011276
24	24	88.9	101	22	AA042882
25	24	88.9	106	11	AA064446
26	24	88.9	106	11	AA091523
27	24	88.9	109	21	AA082577
28	24	88.9	125	20	AA037409
29	24	88.9	139	22	AA026448
30	24	88.9	143	22	AA016447
31	24	88.9	144	22	AA091826
32	24	88.9	144	22	AA080201
33	24	88.9	146	22	AA067650
34	24	88.9	156	23	AA049768
35	24	88.9	186	22	AA027787
36	24	88.9	214	22	AA000371
37	24	88.9	232	22	AA091725
38	24	88.9	259	13	AA024257
39	24	88.9	260	22	AA053799
40	24	88.9	276	21	AA052295
41	24	88.9	296	22	AA098722
42	24	88.9	298	22	AA094028
43	24	88.9	298	22	AA064403
44	24	88.9	298	22	AA094641
45	24	88.9	298	23	AA021764

ALIGNMENTS

RESULT 1

ID	AA08139	standard; peptide; 6 AA.
XX	AA08139:	
AC	04-DEC-2000	(first entry)
DT		
XX		
DE		Peptide modulating activity of heparin, and other glycans.
XX		
KW		Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW		cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW		cartilage differentiation; wound healing.
XX		
OS		Synthetic.
XX		
FT	key	Location/Qualifiers
FT	Misc-difference 1..6	/note="this peptide may be repeated an
FT		unspecified number of times"
PN	WO200045831-A1.	
XX		
PD	10-AUG-2000.	
XX		
PF	02-FEB-2000; 2000WO-US02853.	
XX		
PR	02-FEB-1999; 99US-0118276.	
XX		
PA	(UYJE-) UNIV JEFFERSON THOMAS.	
XX		
PI	San Antonio JD, Verrecchio A, Schick BP;	
XX		

DR WPI: 2000-543446/49.
XX
XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
PS Disclosure: Page 23; 76pp; English.
XX
CC The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
XX
SQ Sequence 6 AA:
Query Match 100.0%; Score 27; DB 21; Length 6;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKKARA 6
DB 1 AKKARA 6
RESULT 2
AAB08180
ID AAB08180 standard; peptide: 12 AA.
XX
AC AAB08180;
XX
DT 04-DEC-2000 (first entry)
XX
DE Peptide modulating activity of heparin, and other glycans.
XX
XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW cartilage differentiation; wound healing.
XX
OS Synthetic.
XX
PN WO200045831-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-US02853.
XX
PR 02-FEB-1999; 99US-0118276.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI San Antonio JD, Verrecchio A, Schick BP;
XX
XX WPI: 2000-543446/49.
XX
PT Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
PS Disclosure: Page 32; 76pp; English.
XX
CC The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with

CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
XX
SQ Sequence 12 AA:
Query Match 100.0%; Score 27; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKKARA 6
DB 1 AKKARA 6
RESULT 3
AAB08181
ID AAB08181 standard; peptide: 24 AA.
XX
AC AAB08181;
XX
DT 04-DEC-2000 (first entry)
XX
DE Peptide modulating activity of heparin, and other glycans.
XX
XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW cartilage differentiation; wound healing.
XX
OS Synthetic.
XX
PN WO200045831-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-US02853.
XX
PR 02-FEB-1999; 99US-0118276.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI San Antonio JD, Verrecchio A, Schick BP;
XX
XX WPI: 2000-543446/49.
XX
PT Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
PS Disclosure: Page 32; 76pp; English.
XX
CC The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.

```
XX Sequence 24 AA:
SQ
Query Match 100.0%; Score 27; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
   |||||
Db 1 AKKARA 6

RESULT 4
AAB08167
ID AAB08167 standard; peptide: 30 AA.
XX
XX AAB08167;
AC
XX 04-DEC-2000 (first entry)
DT
XX
XX Peptide modulating activity of heparin, and other glycans.
DE
XX
XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW cartilage differentiation; wound healing.
XX
XX Synthetic.
OS
XX
XX WO200045831-A1.
PN
XX
XX 10-AUG-2000.
PD
XX
XX 02-FEB-2000; 2000WO-US02853.
PF
XX
XX 02-FEB-1999; 99US-0118276.
PR
XX
XX (UYJE-) UNIV JEFFERSON THOMAS.
PA
XX
XX San Antonio JD, Verrecchio A, Schlick BP;
PI
XX
XX WPI; 2000-543446/49.
DR
XX
XX
XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
XX Disclosure: Page 30; 76pp; English.
PS
XX
XX The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
XX
XX
SQ Sequence 30 AA:
Query Match 100.0%; Score 27; DB 21; Length 30;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
   |||||
Db 1 AKKARA 6
```

```
RESULT 5
AAB08169
ID AAB08169 standard; peptide: 36 AA.
XX
XX AAB08169;
AC
XX 04-DEC-2000 (first entry)
DT
XX
XX Peptide modulating activity of heparin, and other glycans.
DE
XX
XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW cartilage differentiation; wound healing.
XX
XX Synthetic.
OS
XX
XX WO200045831-A1.
PN
XX
XX 10-AUG-2000.
PD
XX
XX 02-FEB-2000; 2000WO-US02853.
PF
XX
XX 02-FEB-1999; 99US-0118276.
PR
XX
XX (UYJE-) UNIV JEFFERSON THOMAS.
PA
XX
XX San Antonio JD, Verrecchio A, Schlick BP;
PI
XX
XX WPI; 2000-543446/49.
DR
XX
XX
XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
XX Disclosure: Page 30; 76pp; English.
PS
XX
XX The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
XX
XX
SQ Sequence 36 AA:
Query Match 100.0%; Score 27; DB 21; Length 36;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
   |||||
Db 1 AKKARA 6

RESULT 6
AAM89334
ID AAM89334 standard; Protein: 89 AA.
XX
XX AAM89334;
AC
XX 07-NOV-2001 (first entry)
DT
XX
XX Human Immune/Haematopoietic antigen SEQ ID NO:16927.
DE
XX
```

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX Homo sapiens.
XX OS
XX XX
PN WO200157182-A2.
XX PD
XX 09-AUG-2001.
XX PF
XX 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001. 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-483426/52.
XX
DR N-PSDB; AAK62115.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
PS Claim 11: SEQ ID NO 16927; 3071bp + Sequence Listing; English.
XX
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM621170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent, CC
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells, especially
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK534942 to AAK54950 and AAM62169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 89, AA;

Query Match	100.0%	Score 27;	DB 22;	Length 89;
Best Local Similarity	100.0%	Pred. No. 81;		
Matches	6; Conservative	0;	Mismatches	0; Gaps

QY	1	AKKARA	6
Db	52	AKKARA	57

RESULT 7
AAV87548
ID AAV87548 standard; Protein; 146 AA

AA
AC AAY87548;

DT 18-JUL-2000 (first entry)

Human disease-associated calmodulin protein (DACP-1).

Disease-associated calmodulin protein; DACP-1; human; EF-hand; KW

reproductive disorder; gynaecological; gene therapy.

OS Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

Accession	Source	Notes
U000000	GenBank	Phosphorylated by casein kinase II

FT /note= "Phosphorylated by casein kinase II"

FT	EF-hand domain	/note=
FT		

FT /note= "N-glycosylated"

FT	/note= "Phosphorylated by protein kinase C"

FT	note= "N-glycosylated"
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

FT	Modified-site	54	"Phosphorylated by casein kinase II"
FT		/note=	
FT	Modified-site	58	"Phosphorylated by casein kinase II"
FT		/note=	
FT	Domain	57..68	"EF-hand domain"
FT		/note=	
FT	Modified-site	65	"Phosphorylated by casein kinase II"
FT		/note=	
FT	Domain	91..102	"EF-hand domain"
FT		/note=	
FT	Modified-site	99	"Phosphorylated by casein kinase II"
FT		/note=	
FT	Domain	127..138	"EF-hand domain"
FT		/note=	

PN	US6046315-A.
XX	
PD	04-APR-2000.

AA 03-NOV-1997; 97US-0963409
PF

PR 03-NOV-1997; 97US-0963409.

PA (INCY -) INCYTE PHARM INC.

PI Hillman J, Corley NC, Shah P;

DR WPT; 2000-282708/24

XX

PT	New polynucleotide encoding a disease associated calmodulin protein
PT	(DACP-1) useful for diagnosing, preventing and treating cancer, immune disorders and reproductive disorders -
PT	

PS Claim 1; Fig 1A-B; 27pp; English.

This sequence represents human disease-associated calmodulin protein (DACP-1). cDNA sequences encoding DACP-1 were initially isolated from a breast tumour cDNA library, the cDNA encoding this sequence being a consensus. Human DACP-1 has four EF-hand calcium-binding domains, and has chemical and structural homology with other calmodulin proteins from human, rat and plasmodium falciparum. DACP-1 proteins and nucleotides are useful for the diagnosis, prevention, or treatment of cancers, immune disorders (e.g., AIDS) and reproductive disorders, including endometriosis. DACP-1 nucleotides are useful for detecting and quantitating gene expression in biopsied tissues in which expression of the protein may be correlated with a disease. Such a diagnostic assay may be used to distinguish between absence, presence, and overexpression of DACP-1, and to monitor regulation of DACP-1 levels during therapeutic intervention. The nucleotide sequences are also useful in assays used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies. In clinical trials, or in monitoring the treatment of an individual patient, Nucleotides encoding DACP-1 are also useful in modulating DACP-1 expression e.g., in antisense therapy. DACP-1, and DACP-1 fragments and antibodies may be used in drug screening techniques. The antibodies may also be useful for the diagnosis of diseases associated with abnormal DACP-1 expression or activity.

AA	Sequence	146 AA;
50		

Query Match	100.0%	Score 27	DB 21	Length 146
Best Local Similarity	100.0%	Pred. NO.	1.3e+02	
Matches 6	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY 1 AKKARA 6

Db 73 AKKARA 78

ABB91739

XX
CT

AC ABB91739;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 950.
XX
KM Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 950; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
CC
SQ Sequence 156 AA;

Query Match 100.0%; Score 27; DB 23; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
Db 144 AKKARA 149

RESULT 9
AAU87578
ID AAU87578 standard; Protein; 163 AA.
XX
AC AAU87578;
XX
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #488.
XX
KM Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KM hyperproliferative disorder; neoplasm; cardiovascular disorder;
KM cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;
KM nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KM acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KM adenocarcinoma; reproductive system disorder; testicular feminisation;
KM endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KM respiratory disorder; renal disorder; kidney failure; blood disorder;
KM myocardial infarction; wound healing; cell proliferation; skin aging;
KM food additive; food preservative; gene therapy.
XX

OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241825.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02359678.
PA XX
XX

(HUMA-)

HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;
PI WPI: 2001-581633/65.
XX N-PSDB: ABK43908.
DR
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
PS Claim 9; SEQ ID NO 1096; 837pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (II) encoded
CC by (I) are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis. Infections caused by bacteria, viruses
CC e.g. Acquired Immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC
Query Match 100.0%; Score 27; DB 22; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKKARA 6
Db 90 AKKARA 95
RESULT 10
AAM43627
ID AAM43627 standard; Protein; 163 AA.
XX
AC AAM43627;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 305.
XX
XX Human: antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antinflammatory; anticancer; vulnary; anticonvulsant; antibacterial;
KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.
OS Homo sapiens.
XX
XX WO200155308-A2.
PN
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01309.
PF
XX 31-JAN-2000; 2000US-0179065.
PR
PR 04-FEB-2000; 2000US-0180628.

PR	24-FEB-2000	2000US-0184664
PR	02-MAR-2000	2000US-0186530
PR	15-MAR-2000	2000US-0189674
PR	17-MAR-2000	2000US-0190076
PR	18-APR-2000	2000US-0198123
PR	19-MAY-2000	2000US-0205515
PR	07-JUN-2000	2000US-0209467
PR	28-JUN-2000	2000US-0214886
PR	30-JUN-2000	2000US-0215135
PR	07-JUL-2000	2000US-0216647
PR	11-JUL-2000	2000US-0216887
PR	11-JUL-2000	2000US-0217496
PR	16-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225568
PR	14-AUG-2000	2000US-0225570
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226275
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226682
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227009
PR	30-ANG-2000	2000US-0229824
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	06-SEP-2000	2000US-0231142
PR	08-SEP-2000	2000US-0231143
PR	08-SEP-2000	2000US-0231144
PR	08-SEP-2000	2000US-0231145
PR	08-SEP-2000	2000US-0231280
PR	08-SEP-2000	2000US-0232064
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0232197
PR	14-SEP-2000	2000US-0232365
PR	14-SEP-2000	2000US-0232366
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234724
PR	25-SEP-2000	2000US-0234974
PR	25-SEP-2000	2000US-0234984
PR	25-SEP-2000	2000US-0234984
PR	26-SEP-2000	2000US-0235488
PR	27-SEP-2000	2000US-0235634
PR	27-SEP-2000	2000US-0235636
PR	29-SEP-2000	2000US-0236127
PR	29-SEP-2000	2000US-0236157
PR	29-SEP-2000	2000US-0236158
PR	29-SEP-2000	2000US-0236159
PR	29-SEP-2000	2000US-0236570
PR	02-OCT-2000	2000US-0236602
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237038

PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241271.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249246.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254057.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	Rosen CA, Barash SC, Ruben SM,	
XX	WPI: 2001-486781/53.	
PL	N-PSDB; AAI6933.	
DR	New isolated nucleic acids and f	
XX	treating and/or preventing humam	
PT		
TS	Claim 11; SEQ ID NO 305; 66pdpp +	

XX The invention relates to human polynucleotides (AA163803-AA164012) and
CC the encoded proteins (AA434497-AA43660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
CC
XX

SO Sequence 163 AA;

Query Match 100.0%; Score 27; DB 22; Length 163;
Best Local Similarity 100.0%; Pred. No. 1,4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
| | | | |
Db 90 AKKARA 95

RESULT 11
AAU19929
ID AAU19929 standard; Protein: 163 AA.

AC AAU19929;

DT 06-DEC-2001 (first entry)

XX Novel human calcium-binding protein #38.

XX Human; calcium-binding protein; calcium flux; neurological disease;

KW Immune dysfunction; digestive disorder; neoplastic disease;

KW blood disorder; infectious disease; gene therapy; immunosuppressive;

KW antiarthritic; cytostatic; vasotropic; antibacterial; nootropic;

KW virucide.

XX Homo sapiens.

OS WO200155304-A2.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01302.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226779.
PR 22-AUG-2000; 2000US-0226861.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231415.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0244676.


```

RESULT 13
AAW49866
ID AAW49866 standard; Protein; 537 AA.
XX
AC AAW49866;
XX
DT 21-DEC-1998 (first entry)
XX
DE Bankia gouldi endoglucanase 37GPI.
XX
KW Glycosidase; 37GPI; thermostable enzyme; oligosaccharide;
KW glucose; sugar; baking; textile; detergent; endoglucanase.
XX
OS Bankia gouldi.
XX
PN WO982479-A1.
XX
PD 11-JUN-1998.
XX
PF 08-DEC-1997; 97WO-US22623.
XX
PR 10-OCT-1997; 97US-0949026.
PR 06-DEC-1996; 96US-0056916.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
XX
DR WPI; 1998-362407/31.
DR N-PSDB; AAW36915.
XX
PT Glycosidase enzymes from organisms of the genera Staphylothermus,
PT Pyrococcus and Thermococcus - for deriving sugar from
PT oligosaccharides, useful in the e.g. food processing, textile or
PT baking industries
XX
PS Claim 1; Fig 9a-c; 92pp; English.
XX
CC This is the amino acid sequence of endoglucanase 37GPI deduced from
CC a polynucleotide (see AAW36915) of Bankia gouldi. The invention
CC provides 18 polynucleotides (see AAW36907-24) coding for thermostable
CC glycosidases (see AAW49858-75) having glucosidase, alpha-galactosidase,
CC beta-galactosidase, beta-mannosidase, beta-mannanase, endoglucanase
CC or pullulanase activity. Vectors and host cells are also claimed.
CC A method is provided for producing the enzymes by recombinant
CC techniques. A claimed method for generating glucose from soluble
CC cell oligosaccharides comprises contacting a sample (selected from
CC dairy products, fruit juice, detergent, textile, guar gum, animal
CC feed, plant biomass or waste product) containing oligosaccharides
CC (selected from maltose, cellobiose, lactose, sucrose, raffinose,
CC stachyose, verbascone, cellulose, starch, amylose, glycogen,
CC disaccharides, polysaccharides and pullulan) with one of the
CC claimed glycosidases such that glucose is produced.
XX
SQ Sequence 537 AA;
XX
Query Match 100.0%; Score 27; DB 19; Length 537;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AKKARA 6
DB 242 AKKARA 247

```

```

XX
DE Pyrococcus abyssi protein #5.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN WO200065062-A2.
XX
PD 02-NOV-2000.
XX
PF 21-APR-2000; 2000WO-FR01065.
XX
PR 21-APR-1999; 99FR-0005034.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX (FRRE-) INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry J, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R, Flament D;
XX Raffin J, Henneke G, Gueguen Y, Rolland J;
XX
DR WPI; 2001-126236/15.
XX
PT New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
XX
PS Disclosure; SEQ ID NO 809; 1403pp; French.
XX
CC The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as FR2792651 but this
CC sequence is not shown in FR2792651, and was obtained in electronic format
CC directly from the European Patent Office.
XX
SQ Sequence 1270 AA;
XX
Query Match 100.0%; Score 27; DB 22; Length 1270;
Best Local Similarity 100.0%; Pred. No. 9.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AKKARA 6
DB 25 AKKARA 30

```

```

RESULT 15
AAV52028
ID AAV52028 standard; Protein; 1434 AA.
XX
AC AAV52028;
XX
DT 12-JUL-2000 (first entry)
XX
DE P. horikoshi PHB021 protein.
XX
KW Polymerase; thermostable; nucleic acid elongation; sliding clamp protein;
KW amplification; reverse transcription.
XX
OS Pyrococcus horikoshii.
XX
PN DEL9840771-A1.
XX
PD 10-FEB-2000.
XX
PF 07-SEP-1998; 98DE-1040771.
XX
PR 06-AUG-1998; 98DE-1035653.

```

XX (LION-) LION BIOSCIENCE AG.
PA
XX
PI Voss H, Moeckel G, Kober I, Kilger C;
XX
DR WPI; 2000-207149/19.
XX
PT A thermostable in vitro polymerase complex for template-dependent
PT elongation of nucleic acids in amplification or reverse transcription
PT methods -
XX
PS Disclosure; Page 100-105; 152pp; German.
XX
CC This invention describes a novel thermostable in vitro complex for
CC template-dependent elongation of nucleic acids which comprises a
CC thermostable sliding clamp protein, which is connected with an
CC elongation protein that shows thermostable polymerase activity. The
CC thermostable in vitro accessory complex can be used to produce the
CC thermostable in vitro complex, which is useful for template-dependent
CC elongation of nucleic acids, e.g. for amplification or reverse
CC transcription. This is useful for sequencing nucleic acids by the
CC polymerase chain reaction or reverse transcriptase-polymerase chain
CC reaction (RT-PCR). The complex can be used to mark nucleic acids.
CC AAY52000-Y52084 and AAW90752-W90799 represent proteins and protein
CC fragments used to illustrate the method of the invention.
XX
SQ Sequence 1434 AA;

Query Match 100.0%; Score 27; DB 21; Length 1434;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
Db 28 AKKARA 33

Search completed: April 8, 2003, 11:23:42
Job time : 27.3448 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:19:38 ; Search time 7.86207 Seconds
(without alignments)
22.454 Million cell updates/sec

Title: US-09-496-391-4
Perfect score: 27
Sequence: 1 AKKARA 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	146	3	US-08-963-409-1
2	27	100.0	537	4	US-09-134-078-23
3	24	88.9	276	2	US-08-989-386-3
4	24	88.9	296	4	US-08-944-604-20
5	24	88.9	301	4	US-08-944-604-18
6	24	88.9	335	4	US-09-134-001C-5021
7	24	88.9	346	4	US-09-352-990-23
8	24	88.9	1044	2	US-08-777-405A-2
9	24	88.9	1044	2	US-08-977-871A-2
10	24	88.9	1044	2	US-09-225-951-2
11	24	88.9	1263	4	US-09-446-504-6
12	24	88.9	1263	4	US-09-712-266-6
13	24	88.9	1263	4	US-09-091-889A-4
14	23	85.2	45	2	US-08-805-191-3
15	23	85.2	59	4	US-09-324-541-2
16	23	85.2	109	4	US-08-905-223-313
17	23	85.2	231	4	US-09-370-098-4
18	23	85.2	300	4	US-09-254-465A-10
19	23	85.2	381	4	US-09-183-861-81
20	23	85.2	381	4	US-09-022-765-81
21	23	85.2	396	4	US-09-142-551A-3
22	23	85.2	404	3	US-09-046-578-2
23	23	85.2	404	3	US-09-046-578-4
24	23	85.2	404	3	US-09-046-578-6
25	23	85.2	417	4	US-08-887-534A-38
26	23	85.2	417	4	US-08-887-534A-40
27	23	85.2	417	4	US-08-887-534A-42

28	23	85.2	439	4	US-08-975-762-60	Sequence 60, Appl
29	23	85.2	439	4	US-09-295-028-60	Sequence 60, Appl
30	23	85.2	439	4	US-09-106-582-60	Sequence 60, Appl
31	23	85.2	474	4	US-09-461-474-10	Sequence 10, Appl
32	23	85.2	516	4	US-08-762-106-8	Sequence 8, Appl
33	23	85.2	516	3	US-08-745-404-2	Sequence 2, Appl
34	23	85.2	516	4	US-09-320-774-8	Sequence 8, Appl
35	23	85.2	527	2	US-08-762-106-9	Sequence 9, Appl
36	23	85.2	527	4	US-09-320-774-9	Sequence 9, Appl
37	23	85.2	552	3	US-08-745-404-3	Sequence 3, Appl
38	23	85.2	606	4	US-09-362-123A-6	Sequence 6, Appl
39	23	85.2	610	2	US-08-879-561-10	Sequence 10, Appl
40	23	85.2	624	2	US-08-879-561-3	Sequence 3, Appl
41	23	85.2	624	2	US-08-879-561-9	Sequence 9, Appl
42	23	85.2	648	4	US-09-134-001C-5161	Sequence 5161, Ap
43	23	85.2	705	4	US-09-370-838-186	Sequence 186, App
44	23	85.2	746	5	PCT-US95-10509-2	Sequence 2, Appl
45	23	85.2	826	4	US-09-564-805-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-963-409-1
Sequence 1, Application US/08963409

Patent No. 6046315

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Shah, Purvi

APPLICANT: Corley, Neil C.

TITLE OF INVENTION: DISEASE ASSOCIATED CALMODULIN PROTEIN

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESS: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/963,409

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0418 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 146 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: BRSTTUT14

CLONE: 2743380

US-08-963-409-1

Query Match 100.0%; Score 27; DB 3; Length 146;
Best local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
Db 73 AKKARA 78

RESULT 2

US-09-134-078-23
; Sequence 23, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ. ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-09-134-078-23

Query Match 100.0%; Score 27; DB 4; Length 537;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
Db 242 AKKARA 247

RESULT 3
US-08-989-386-3
; Sequence 3, Application US/08989386
; Patent No. 5989860
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Puri
; TITLE OF INVENTION: HUMAN ISOMERASE HOMOLOGS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/989,386
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0443 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166

INFORMATION FOR SEQ. ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 276 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:
LIBRARY: TMLR3DPT01
CLONE: 292808

US-08-989-386-3

Query Match 88.9%; Score 24; DB 2; Length 276;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
Db 133 AKKARS 138

RESULT 4
US-08-944-604-20
; Sequence 20, Application US/08944604
; Patent No. 6218131
; GENERAL INFORMATION:
; APPLICANT: KEESER, SUSAN
; APPLICANT: OBAR, ROBERT
; APPLICANT: WU, YING-JYE

TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Testa, Hurwitz & Thibault
STREET: 125 High St.
CITY: Boston
STATE: MA
COUNTRY: USA

ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,604
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: MEYERS, THOMAS C
REGISTRATION NUMBER: 36,989
REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-944-604-20

Query Match 88.9%; Score 24; DB 4; Length 296;
Best Local Similarity 83.3%; Pred. No. 4.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
|||||
Db 133 AKKARS 138

RESULT 5
US-08-944-604-18
Sequence 18, Application US/08944604
Patent No. 6218131
GENERAL INFORMATION:
APPLICANT: KEESEE, SUSAN
APPLICANT: OBAR, ROBERT
APPLICANT: WU, YING-JYE
TITLE OF INVENTION: MATERIALS AND METHODS FOR DETECTION OF
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Testa, Hurwitz & Thibault
STREET: 125 High St.
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,604
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MEYERS, THOMAS C
REGISTRATION NUMBER: 36,989
REFERENCE/DOCKET NUMBER: MTP-021 (8395/24)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 301 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-944-604-18

Query Match 88.9%; Score 24; DB 4; Length 301;
Best Local Similarity 83.3%; Pred. No. 4.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
|||||
Db 133 AKKARS 138

RESULT 6
US-09-134-001C-5021
Sequence 5021, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5021
LENGTH: 335
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5021

Query Match 88.9%; Score 24; DB 4; Length 335;
Best Local Similarity 83.3%; Pred. No. 5.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
|||||
Db 179 AKKARA 184

RESULT 7
US-09-352-990-23
Sequence 23, Application US/09352990
Patent No. 6255090
GENERAL INFORMATION:
APPLICANT: Famodu, Layo O.
APPLICANT: Orozco, Buddy
APPLICANT: Rafalski, Antoni
TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
FILE REFERENCE: BB-1191
CURRENT APPLICATION NUMBER: US/09/352,990
CURRENT FILING DATE: 1999-07-14
EARLIER APPLICATION NUMBER: 60/092,866
EARLIER FILING DATE: July 15, 1998
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Microsoft Office 97
SEQ ID NO 23
LENGTH: 346
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-352-990-23

Query Match 88.9%; Score 24; DB 4; Length 346;
Best Local Similarity 83.3%; Pred. No. 5.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
|||||
Db 14 AKKARA 19

RESULT 8
US-08-777-405A-2
Sequence 2, Application US/08777405A
Patent No. 5858753
GENERAL INFORMATION:
APPLICANT: Chantry, David
APPLICANT: Hoekstra, Merl F.
APPLICANT: Holtzman, Douglas A
TITLE OF INVENTION: No. 5858753e1 Lipid Kinase
NUMBER OF SEQUENCES: 17

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall O'Toole Gerstein Murray & Borun
;; STREET: 6300 Sears Tower/233 South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60606
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentln Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/777,405A
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: No. 5858753and, Greta E.
;; REGISTRATION NUMBER: 35,302
;; REFERENCE/DOCKET NUMBER: 27866/33441
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312) 474-6300
;; TELEFAX: (312) 474-0448
;; TELEX: 25-3856
;;
;; INFORMATION FOR SEQ. ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1044 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
;; US-08-777-405A-2
;;
Query Match 88.9%; Score 24; DB 2; Length 1044;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKKARA 6
Db 401 AKKARS 406
;;
RESULT 9
;; US-08-977-871A-2
;; Sequence 2, Application US/08977871A
;; Patent No. 5882910
;; GENERAL INFORMATION:
;; APPLICANT: Chantry, David
;; APPLICANT: Hoekstra, Merl F.
;; APPLICANT: Holtzman, Douglas A
;; TITLE OF INVENTION: No. 5882910el Lipid Kinase
;; NUMBER OF SEQUENCES: 17
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall O'Toole Gerstein Murray & Borun
;; STREET: 6300 Sears Tower/233 South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentln Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/977,871A
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/777,405
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: No. 5882910and, Greta E.
;; REGISTRATION NUMBER: 35,302

;; REFERENCE/DOCKET NUMBER: 27866/33441
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312) 474-6300
;; TELEFAX: (312) 474-0448
;; TELEX: 25-3856
;;
;; INFORMATION FOR SEQ. ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1044 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
;; US-08-977-871A-2
;;
Query Match 88.9%; Score 24; DB 2; Length 1044;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKKARA 6
Db 401 AKKARS 406
;;
RESULT 10
;; US-09-225-951-2
;; Sequence 2, Application US/09225951
;; Patent No. 5985589
;; GENERAL INFORMATION:
;; APPLICANT: Chantry, David
;; APPLICANT: Hoekstra, Merl F.
;; APPLICANT: Holtzman, Douglas A
;; TITLE OF INVENTION: No. 5985589el Lipid Kinase
;; NUMBER OF SEQUENCES: 17
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Marshall O'Toole Gerstein Murray & Borun
;; STREET: 6300 Sears Tower/233 South Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentln Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/225,951
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: No. 5985589and, Greta E.
;; REGISTRATION NUMBER: 35,302
;; REFERENCE/DOCKET NUMBER: 27866/33441
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312) 474-6300
;; TELEFAX: (312) 474-0448
;; TELEX: 25-3856
;;
;; INFORMATION FOR SEQ. ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1044 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
;; US-09-225-951-2
;;
Query Match 88.9%; Score 24; DB 2; Length 1044;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKKARA 6
Db 401 AKKARS 406

RESULT 11
US-09-446-504-6
; Sequence 6, Application US/09446504
; Patent No. 6218150
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshimi
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazue
; APPLICANT: MUKAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/446,504
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1263
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-446-504-6

Query Match 88.9%; Score 24; DB 4; Length 1263;
Best Local Similarity 83.3%; Pred. No. 1.6e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
Db 25 AKKARS 30

RESULT 12
US-09-712-266-6
; Sequence 6, Application US/09712266
; Patent No. 6333158
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshimi
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazue
; APPLICANT: MUKAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/712,266
; CURRENT FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 09/446,504
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1263
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-712-266-6

Query Match 88.9%; Score 24; DB 4; Length 1263;
Best Local Similarity 83.3%; Pred. No. 1.6e+03;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 AKKARA 6
Db 25 AKKARS 30

RESULT 13
US-09-091-889A-4
; Sequence 4, Application US/09091889A
; Patent No. 6395526
; GENERAL INFORMATION:
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: ISHINO, YOSHIZUMI
; APPLICANT: KATO, IKUNOSHIN
; TITLE OF INVENTION: NOVEL DNA POLYMERASE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,889A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, MARC S.
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 1422-0347P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1263 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-091-889A-4

Query Match 88.9%; Score 24; DB 4; Length 1263;
Best Local Similarity 83.3%; Pred. No. 1.6e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKARA 6
Db 25 AKKARS 30

RESULT 14
US-08-805-191-3
; Sequence 3, Application US/08805191
; Patent No. 5908924
; GENERAL INFORMATION:
; APPLICANT: Burdette, Douglas S
; APPLICANT: Zeikus, Joseph G
; TITLE OF INVENTION: CLONING AND EXPRESSION OF THE GENE
; TITLE OF INVENTION: ENCODING THERMOANEROBACTER ETHANOLICUS 39E
; TITLE OF INVENTION: SECONDARY-ALCOHOL DEHYDROGENASE AND ENZYME BIOCHEMICAL
; TITLE OF INVENTION: CHARACTERIZATION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pluckney Street
; CITY: Madison

STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,191
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/012,331
FILING DATE: 27-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berson, Bennett J
REGISTRATION NUMBER: 37094
REFERENCE/DOCKET NUMBER: 660336, 90608
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-805-191-3

Query Match 85.2%; Score 23; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKKAR 5
DB 25 AKKAR 29

RESULT 15
US-09-324-541-2
; Sequence 2, Application US/09324541
; Patent No. 6391855
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: JUNCTIONAL ADHESION MOLECULE-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086, 412
; CURRENT APPLICATION NUMBER: US/09/324,541
; CURRENT FILING DATE: 1999-06-02
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-324-541-2

Query Match 85.2%; Score 23; DB 4; Length 59;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKKARA 6
DB 25 AKKTRA 30

Search completed: April 8, 2003, 11:31:12
Job time: 9.86207 secs

GenCore version 5.1.4.p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:28:39 ; Search time 8.27586 Seconds
(without alignments)
44.324 Million cell updates/sec

Title: US-09-496-391-4

Perfect score: 27

Sequence: 1 AKKARA 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	537	9 US-10-121-032-23	Sequence 23, Appl
2	24	88.9	45	10 US-09-816-989A-2	Sequence 2, Appl
3	24	88.9	66	10 US-09-816-989A-4	Sequence 4, Appl
4	24	88.9	106	9 US-09-738-626-5277	Sequence 5277, Ap
5	24	88.9	109	10 US-09-816-989A-7	Sequence 7, Appl
6	24	88.9	144	9 US-09-738-626-5580	Sequence 5580, Ap
7	24	88.9	232	9 US-09-738-626-5479	Sequence 5479, Ap
8	24	88.9	298	10 US-09-848-852A-1	Sequence 1, Appl
9	24	88.9	1044	9 US-10-027-591-2	Sequence 2, Appl
10	24	88.9	1263	10 US-09-871-309-6	Sequence 6, Appl
11	23	85.2	6	9 US-09-823-823-68	Sequence 68, Appl
12	23	85.2	6	10 US-09-823-829-68	Sequence 68, Appl
13	23	85.2	59	9 US-10-119-537-2	Sequence 1130, Ap
14	23	85.2	87	10 US-09-764-864-1130	Sequence 1130, Ap
15	23	85.2	89	10 US-09-764-864-1545	Sequence 1545, Ap
16	23	85.2	112	10 US-09-764-860-333	Sequence 333, App
17	23	85.2	162	9 US-10-156-239-40	Sequence 40, Appl
18	23	85.2	162	10 US-09-795-693-40	Sequence 40, Appl
19	23	85.2	182	10 US-09-216-393-37	Sequence 37, Appl

20	23	85.2	193	10 US-09-841-132-581	Sequence 581, App
21	23	85.2	197	10 US-09-925-300-1178	Sequence 1178, Ap
22	23	85.2	201	10 US-09-741-669-457	Sequence 457, App
23	23	85.2	223	12 US-10-062-254-204	Sequence 204, App
24	23	85.2	231	10 US-09-878-486-4	Sequence 4, Appl
25	23	85.2	243	10 US-09-771-161A-127	Sequence 127, App
26	23	85.2	260	9 US-10-102-806-503	Sequence 503, App
27	23	85.2	260	9 US-09-925-300-1470	Sequence 1470, Ap
28	23	85.2	300	9 US-09-953-499-10	Sequence 10, Appl
29	23	85.2	334	10 US-09-815-242-11070	Sequence 11070, A
30	23	85.2	380	9 US-09-981-496-106	Sequence 106, App
31	23	85.2	380	10 US-09-874-923-106	Sequence 106, App
32	23	85.2	381	9 US-09-991-496-81	Sequence 81, Appl
33	23	85.2	381	10 US-09-874-923-81	Sequence 81, Appl
34	23	85.2	439	10 US-09-159-469-60	Sequence 60, Appl
35	23	85.2	439	10 US-09-798-042-60	Sequence 60, Appl
36	23	85.2	484	9 US-09-738-626-6430	Sequence 6430, App
37	23	85.2	487	10 US-09-815-242-5183	Sequence 5183, Ap
38	23	85.2	529	10 US-09-815-242-11258	Sequence 11258, A
39	23	85.2	559	9 US-09-866-050A-499	Sequence 499, App
40	23	85.2	593	9 US-09-738-626-5063	Sequence 5063, Ap
41	23	85.2	624	9 US-09-984-245-196	Sequence 196, App
42	23	85.2	624	9 US-09-986-262-196	Sequence 196, App
43	23	85.2	624	9 US-09-983-466-196	Sequence 196, App
44	23	85.2	624	10 US-09-771-161A-244	Sequence 244, App
45	23	85.2	644	10 US-09-815-242-5670	Sequence 5670, Ap

ALIGNMENTS

RESULT 1
US-10-121-032-23
Sequence 23, Application US/10121032
Patent No. US20020155550A1
GENERAL INFORMATION:
APPLICANT: BYLINA, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/121.032
FILING DATE: 09-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/134.078
FILING DATE: 13-AUG-1998
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858/677-1456
TELEFAX: 858/677-1465
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-121-032-23

Query Match      100.0%; Score 27; DB 9; Length 537;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AKKARA 6
       111111
Db      242 AKKARA 247

RESULT 2
US-09-816-989A-2
; Sequence 2, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-2

Query Match      88.9%; Score 24; DB 10; Length 45;
Best Local Similarity 83.3%; Pred. No. 59;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AKKARA 6
       111111
Db      5 AKKARA 10

RESULT 3
US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

```

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-4

Query Match      88.9%; Score 24; DB 10; Length 66;
Best Local Similarity 83.3%; Pred. No. 86;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AKKARA 6
       111111
Db      22 AKKARA 27

RESULT 4
US-09-738-626-5277
; Sequence 5277, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5277
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5277

Query Match      88.9%; Score 24; DB 9; Length 106;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 AKKARA 6
       111111
Db      21 AKKARA 26

RESULT 5
US-09-816-989A-7
; Sequence 7, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lis, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT M
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
```

LENGTH: 109
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-7

Query Match 88.9%; Score 24; DB 10; Length 109;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
111111
DB 13 AKKARA 18

RESULT 6
US-09-738-626-5580
Sequence 5580, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5580
LENGTH: 144
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5580

Query Match 88.9%; Score 24; DB 9; Length 144;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
111111
DB 79 AKKARA 84

RESULT 7
US-09-738-626-5479
Sequence 5479, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5479
LENGTH: 232
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5479

Query Match 88.9%; Score 24; DB 9; Length 232;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
111111
DB 167 AKKARA 172

RESULT 8
US-09-848-852A-1
Sequence 1, Application US/09848852A
Patent No. US20020106373A1
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
Tang, Y. Tom
Corley, Neil C.
Guebler, Karl J.
Yue, Henry
Patterson, Chandra
TITLE OF INVENTION: PROTEINS ASSOCIATED WITH EGF SIGNALING
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESS: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/848, 852A
FILING DATE: 07-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/069, 725
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0515 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 298 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNOT16

CLONE: 1281694
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-848-852A-1

Query Match 88.9%; Score 24; DB 10; Length 298;
Best Local Similarity 83.3%; Pred. No. 3.7e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
|||||:
DB 140 AKKARA 145

RESULT 9
US-10-027-591-2
Sequence 2, Application US/10027591
Patent No. US20020161014A1
GENERAL INFORMATION:
APPLICANT: SADHU, Chanchal et al.
TITLE OF INVENTION: INHIBITORS OF HUMAN PHOSPHATIDYLINOSITOL 3-KINASE DELTA
FILE REFERENCE: 27866/36170C
CURRENT APPLICATION NUMBER: US/10/027,591
CURRENT FILING DATE: 2001-10-19
PRIOR APPLICATION NUMBER: 09/841,341
PRIOR FILING DATE: 2001-04-24
PRIOR APPLICATION NUMBER: 60/199,655
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 60/238,057
PRIOR FILING DATE: 2000-10-25
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin version 3.0
SEQ ID NO 2
LENGTH: 1044
TYPE: PRT
ORGANISM: Human p110delta protein
US-10-027-591-2

Query Match 88.9%; Score 24; DB 9; Length 1044;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
|||||:
DB 401 AKKARS 406

RESULT 10
US-09-971-309-6
Sequence 6, Application US/09971309
Patent No. US20020106675A1
GENERAL INFORMATION:
APPLICANT: UEMORI, Takashi
APPLICANT: SATO, Yoshimi
APPLICANT: FUJITA, Tomoko
APPLICANT: MIYAKE, Kazuo
APPLICANT: MUKAI, Hiroyuki
APPLICANT: ASADA, Kiyozi
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
FILE REFERENCE: 1422-0494P
CURRENT APPLICATION NUMBER: US/09/971,309
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: US 09/446,504
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: PCT/JP98/02845
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: JP 9-187496
PRIOR FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: JP 9-320692
PRIOR FILING DATE: 1997-11-27
NUMBER OF SEQ ID NOS: 92
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6

LENGTH: 1263
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-09-971-309-6

Query Match 88.9%; Score 24; DB 10; Length 1263;
Best Local Similarity 83.3%; Pred. No. 1.5e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
|||||:
DB 25 AKKARS 30

RESULT 11
US-09-823-823-68
Sequence 68, Application US/09823823
Patent No. US20020171092A1
GENERAL INFORMATION:
APPLICANT: Yamamoto, Satoshi
APPLICANT: Kasai, Hiroaki
APPLICANT: Nakamura, Shoko
APPLICANT: Suzuki, Makoto
TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS USING
FILE REFERENCE: 12817-004001
CURRENT APPLICATION NUMBER: US/09/823,823
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/208,688
PRIOR FILING DATE: 1998-12-10
PRIOR APPLICATION NUMBER: JP 97/343316
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 80
SOFTWARE: Patentin version 2.0
SEQ ID NO 68
LENGTH: 6
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetically generated protein
US-09-823-823-68

Query Match 85.2%; Score 23; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKAR 5
|||||:
DB 1 AKKAR 5

RESULT 12
US-09-823-829-68
Sequence 68, Application US/09823829
Patent No. US2002014697A1
GENERAL INFORMATION:
APPLICANT: Yamamoto, Satoshi
APPLICANT: Nakamura, Shoko
APPLICANT: Suzuki, Makoto
APPLICANT: Kasai, Hiroaki
TITLE OF INVENTION: METHOD FOR IDENTIFICATION AND DETECTION OF MICROORGANISMS
FILE REFERENCE: 12817-004001
CURRENT APPLICATION NUMBER: US/09/823,829
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 09/208,688
PRIOR FILING DATE: 1998-12-10
PRIOR APPLICATION NUMBER: JP 97/343316
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 82
SOFTWARE: Patentin version 2.0


```

; SEQ ID NO 68
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetically generated peptide
US-09-823-829-68

```

```

Query Match      85.2%; Score 23; DB 10; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AKKAR 5
      11111
Db      1 AKKAR 5

```

```

RESULT 13
US-10-119-537-2
; Sequence 2, Application US/10119537
; Publication No. US20030027761A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: JUNCTIONAL ADHESION MOLECULE-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086.412D1
; CURRENT APPLICATION NUMBER: US/10/119,537
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-119-537-2

```

```

Query Match      85.2%; Score 23; DB 9; Length 59;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 AKKARA 6
      11111
Db      25 AKKTTRA 30

```

```

RESULT 14
US-09-764-864-1130
; Sequence 1130, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1130
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1130

```

```

Query Match      85.2%; Score 23; DB 10; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AKKAR 5
      11111
Db      55 AKKAR 59

```

```

RESULT 15
US-09-764-864-1545
; Sequence 1545, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1545
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (47)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1545

```

```

Query Match      85.2%; Score 23; DB 10; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AKKAR 5
      11111
Db      57 AKKAR 61

```

```

Search completed: April 8, 2003, 11:52:41
Job time : 10.2759 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:18:53 ; Search time 9.51724 Seconds
(without alignments)
60.606 Million cell updates/sec

Title: US-09-496-391-4

Perfect score: 27

Sequence: 1 AKKARA 6

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	156	2	D84499
2	27	100.0	210	2	A25550
3	27	100.0	651	2	S65296
4	27	100.0	749	2	C72725
5	27	100.0	1434	2	G71232
6	27	100.0	1455	2	E75199
7	24	88.9	70	2	H82602
8	24	88.9	100	2	E72547
9	24	88.9	125	2	A39396
10	24	88.9	125	2	A71477
11	24	88.9	125	2	B81739
12	24	88.9	135	2	S25119
13	24	88.9	141	2	PS0147
14	24	88.9	156	2	T44769
15	24	88.9	156	2	AF1784
16	24	88.9	156	2	AB1408
17	24	88.9	170	2	AB0192
18	24	88.9	171	1	HSUR1E
19	24	88.9	176	2	F88484
20	24	88.9	178	2	T32722
21	24	88.9	186	2	G64486
22	24	88.9	204	2	E64422
23	24	88.9	206	2	S09388
24	24	88.9	218	2	B69934
25	24	88.9	222	2	S33204
26	24	88.9	248	1	HSURIP
27	24	88.9	250	2	S25067
28	24	88.9	259	1	C47069
29	24	88.9	269	2	B98248

30	24	88.9	269	2	A13037	conserved hypothet
31	24	88.9	269	2	JC7700	38k ribosome-assoc
32	24	88.9	278	2	T46458	hypothetical prote
33	24	88.9	296	2	T38492	hypothetical prote
34	24	88.9	301	1	S66681	peptidylprolyl iso
35	24	88.9	322	2	A70607	hypothetical prote
36	24	88.9	334	2	T41705	probable 2-hydroxy
37	24	88.9	334	2	T42743	hypothetical prote
38	24	88.9	345	2	AE0141	biotin synthase (E
39	24	88.9	357	2	C72746	probable carbamoyl
40	24	88.9	441	2	E72242	Mg-protocorphylin
41	24	88.9	444	2	E69893	biotin carboxylase
42	24	88.9	530	2	B70810	hypothetical prote
43	24	88.9	580	2	B70868	probable transfera
44	24	88.9	600	2	S65788	ku antigen 70k cha
45	24	88.9	622	2	T21090	hypothetical prote

ALIGNMENTS

RESULT 1
D84499
hypothetical protein At2g11830 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: D84499
R:Lin. X.: Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, U.A.; Shen, W.; Vanden, S.E.; Umayam, L.; Tallon
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84499
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <STO>
A:Cross-References: GB:AE002093; NID:g4733999; PIDN:AAD28677.1; GSPDB:GND0139
C:Genetics:
A:Gene: At2g11830
A:Map position: 2

Query Match
Best Local Similarity 100.0%; Score 27; DB 2; Length 156;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
DB 144 AKKARA 149

RESULT 2
A25550
histone H1 - sea urchin (lytechinus pictus)
C:Species: Lytechinus pictus (painted urchin)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 23-Jul-1999
A:Accession: A25550
R:Knowles, J.A.; Childs, G.J.
Nucleic Acids Res. 14, 8121-8133, 1986
A:Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus
A:Reference number: A25550; MUID:87040778; PMID:3022245
A:Accession: A25550
A:Molecule type: DNA
A:Residues: 1-210 <KNO>
A:Cross-References: GB:X04488; NID:g9616; PIDN:CAA28177.1; PID:g9617
C:Superfamily: histone H1
C:Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match
Best Local Similarity 100.0%; Score 27; DB 2; Length 210;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6

DB 9 AKKARA 14

RESULT 3

hypothetical protein YPL263c - yeast (*Saccharomyces cerevisiae*)

C:Species: *Pyrococcus horikoshii*

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C:Accession: G71232

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hatakeyama, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusuda, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: G71232

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1434 <KAW>

A:Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29190.1; PID:g3256507

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenB

C:Genetics:

A:Gene: PH0121

Query Match 100.0%; Score 27; DB 2; Length 1434;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6

DB 28 AKKARA 33

RESULT 6

DNA-directed DNA polymerase (EC 2.7.7.7) II large chain PAB2404 - *Pyrococcus abyssi* (

C:Species: *Pyrococcus abyssi*

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: E75199

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome s

A:Reference number: A75001

A:Accession: E75199

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1455 <KAW>

A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49044.1; PID:g545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB2404

C:Keywords: nucleotidyltransferase

Query Match 100.0%; Score 27; DB 2; Length 1455;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6

DB 25 AKKARA 30

RESULT 7

hypothetical protein XP2073 [imported] - *Xylella fastidiosa* (strain 9a5c)

C:Species: *Xylella fastidiosa*

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: H82602

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: H82602

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-70 <SIM>

A:Cross-references: GB:AE004023; GB:AE003849; NID:g9107185; PIDN:AAF84872.1; GSPDB:GN

Query Match 100.0%; Score 27; DB 2; Length 749;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6

DB 48 AKKARA 53

RESULT 5

G71232

A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brines, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H
as-Neto, E.; Docena, C.; El-Dorriy, H.; Falcinani, A.P.; Ferreira, A.J.S.
Submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzima, J.P.; Krieger, J.E.; Kurame, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J.; de Sa, R.G.; Santelli, R.V.; Savasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsubako, M.R.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2073

Query Match 88.9%; Score 24; DB 2; Length 70;
Best Local Similarity 83.3%; Pred. No. 82;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AKKARA 6
Db 22 AKKARA 27

RESULT 8

E72547
hypothetical protein APE1665 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: E72547

R:Kikarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takah
awa, H.; Takamaya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K
DNA Res. 6: 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: E72547

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-100 <KAM>
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BAA80666.1; PID:d1044452; PID:9510
A:Experimental source: strain K1
C:Genetics:

A:Gene: APE1665
Query Match 88.9%; Score 24; DB 2; Length 100;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AKKARA 6
Db 41 SKKARA 46

RESULT 9

A39396
histone H1-like protein Hc1 - Chlamydia trachomatis

C:Species: Chlamydia trachomatis
C>Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 08-Oct-1999
C:Accession: A39396; A41383; S16152

R:Hackstadt, T.; Baehr, W.; Ying, Y.
Proc. Natl. Acad. Sci. U.S.A. 88: 3937-3941, 1991
A:Title: Chlamydia trachomatis developmentally regulated protein is homologous to eukary
A:Reference number: A39396; MUID:91219487; PMID:2023942
A:Accession: A39396

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-125 <HAC>
A:Cross-references: GB:M60902; NID:9144506; PIDN:AAA23129.1; PID:9144507
R:Hackstadt, T.
J. Bacteriol. 173: 7046-7049, 1991

A:Title: Purification and N-terminal amino acid sequences of Chlamydia trachomatis hi
A:Reference number: A41383; MUID:92041597; PMID:1938908
A:Accession: A41383

A:Status: preliminary
A:Molecule type: protein
A:Residues: 2-38 <HA2>

R:Tao, S.; Kaul, R.; Wenman, W.M.
J. Bacteriol. 173: 2818-2822, 1991

A:Title: Identification and nucleotide sequence of a developmentally regulated gene e
A:Reference number: S16152; MUID:91210171; PMID:1708378
A:Accession: S16152

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-125 <TAO>

A:Cross-references: EMBL:X57311; NID:940688; PIDN:CAA40563.1; PID:940689
C:Keywords: DNA binding

Query Match 88.9%; Score 24; DB 2; Length 125;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AKKARA 6
Db 69 AKKARA 74

RESULT 10

A71477
probable histone-like developmental protein - Chlamydia trachomatis (serotype D, stra

C:Species: Chlamydia trachomatis
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: A71477

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282: 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:9900809; PMID:9784136
A:Accession: A71477

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-125 <ARN>

A:Cross-references: GB:AE001345; GB:AE001273; NID:93329198; PIDN:AAC68338.1; PID:9332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:

A:Gene: hcta

Query Match 88.9%; Score 24; DB 2; Length 125;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AKKARA 6
Db 69 AKKARA 74

RESULT 11

B81739
histone H1-like protein Hc1 TC0119 [imported] - Chlamydia muridarum (strain N19g)

C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: B81739

R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28: 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: B81500; MUID:20150255; PMID:10684935
A:Accession: B81739

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-125 <TEH>
A:Cross-references: GB:AE002279; GB:AE002160; NID:97190148; PIDN:AAF38997.1; PID:9719
A:Experimental source: strain N19g (Mopn)
C:Genetics:

A:Gene: TC0119

Query Match 88.9%; Score 24; DB 2; Length 125;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKKARA 6
||||:|
Db 69 AKKARA 74

RESULT 12
S25119
histone H2A - Trypanosoma cruzi
C:Species: Trypanosoma cruzi
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: S25119; S33211
R: Lopez, M.C.
submitted to the EMBL Data Library, July 1992
A:Reference number: S25119
A:Accession: S25119
A:Molecule type: DNA
A:Residues: 1-135 <LOP>
A:Cross-references: EMBL:X67287; NID:g10617; PIDN:CAA47703.1; PID:g10618
C:Superfamily: histone H2A
C:Keywords: chromosomal protein; DNA binding; nucleosome core; nucleus

Query Match 88.9%; Score 24; DB 2; Length 135;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
||||:|
Db 126 SKKARA 131

RESULT 13
PS0147
histone H1 - sea urchin (Strongylocentrotus nudus) (fragment)
C:Species: Strongylocentrotus nudus
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 17-Mar-1994
C:Accession: PS0147
R: Suzuki, M.; Sohma, H.; Yazawa, M.; Yagi, K.; Ebdashi, S.
J. Biochem. 108, 356-364, 1990
A:Title: Histone H1 kinase specific to the SPKK motif.
A:Reference number: PS0147; MUID:91115779; PMID:2177468
A:Accession: PS0147
A:Molecule type: protein
A:Residues: 1-141 <SUZ>
C:Superfamily: histone H1
C:Keywords: chromosomal protein; nucleosome
F:45-116/Region: globular head

Query Match 88.9%; Score 24; DB 2; Length 141;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
||||:|
Db 131 AKKARA 136

RESULT 14
T44769
hypothetical protein [imported] - Bacillus subtilis (fragment)
C:Species: Bacillus subtilis
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000
C:Accession: T44769
R: Moriyama, R.; Hattori, A.; Miyata, S.; Kudoh, S.; Makino, S.
J. Bacteriol. 178, 6059-6063, 1996
A:Title: A gene (slieB) encoding a spore cortex-lytic enzyme from Bacillus subtilis and
A:Reference number: Z22836; MUID:96427343; PMID:8830707
A:Accession: T44769
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-156 <MOR>
A:Cross-references: EMBL:D79978; NID:g1688021; PIDN:BAAL1472.1; PID:g1688022
A:Experimental source: strain 168
C:Superfamily: conserved hypothetical protein ypdC

Query Match 88.9%; Score 24; DB 2; Length 156;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
||||:|
Db 145 AKKARS 150

RESULT 15
AF1784
conserved hypothetical protein 11n2820 [imported] - Listeria innocua (strain Cl1p1126
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF1784
R: Claser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maltournam, A.;
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1784
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC38046.1; PID:g16415356; GSPDB:GN00178
C:Genetics:
A:Experimental source: strain Cl1p11262
A:Gene: 11n2820

Query Match 88.9%; Score 24; DB 2; Length 156;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKKARA 6
||||:|
Db 57 AKKARA 62

Search completed: April 8, 2003, 11:29:50
Job time : 12.5172 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:11:45 ; Search time 63.3621 seconds
(without alignments)
31.545 Million cell updates/sec

Title: US-09-496-391-9

Perfect score: 90

Sequence: 1 YPARRARYQWVRCKP 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database : A.Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	100.0	15	21	AA080141
2	90	100.0	49	14	AA042842
3	77	85.6	15	21	AA080142
4	77	85.6	69	22	AA081015
5	77	85.6	69	22	AA081740
6	77	85.6	69	22	AA081991
7	77	85.6	76	22	AA080982
8	77	85.6	76	22	AA081471
9	77	85.6	76	22	AA081724
10	77	85.6	76	22	AA081954

11	77	85.6	123	22	AA080996
12	77	85.6	123	22	AA081965
13	77	85.6	143	22	AA081008
14	77	85.6	143	22	AA081975
15	77	85.6	148	22	AA081000
16	77	85.6	156	22	AA081993
17	77	85.6	158	11	AA0805247
18	77	85.6	158	11	AA0805247
19	77	85.6	158	14	AA080993
20	77	85.6	158	22	AA080992
21	77	85.6	158	22	AA081961
22	77	85.6	162	21	AA056784
23	75	83.3	164	22	AA030481
24	68	75.6	164	22	AA030481
25	63	70.0	16	22	AA09428
26	61	67.8	201	22	AA031060
27	53	58.9	125	22	AA080937
28	53	58.9	125	22	AA081928
29	47.5	52.8	122	23	ABG65138
30	47.5	52.8	122	23	ABG65139
31	47.5	52.8	122	23	AB094022
32	47.5	52.8	122	23	AB094046
33	47	52.2	494	22	ABG03323
34	45	50.0	415	22	AA052305
35	44	48.9	52	22	AA042873
36	44	48.9	54	21	AA026253
37	44	48.9	85	22	ABG28206
38	44	48.9	95	22	ABG28205
39	44	48.9	99	22	ABG28204
40	44	48.9	117	22	ABG28201
41	42	46.7	53	23	ABG65677
42	42	46.7	207	21	AA044338
43	42	46.7	785	22	AB062933
44	41.5	46.1	53	22	AA053533
45	41	45.6	38	21	AA01942

ALIGNMENTS

RESULT 1	AA080141
ID	AA080141 standard; peptide; 15 AA.
XX	AA080141:
AC	04-DEC-2000 (first entry)
XX	
DT	Serglycin peptide modulating activity of heparin, and other glycans.
XX	
DE	
XX	
KW	Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW	cell attachment; cell adhesion; vein graft; tumour cell metastasis;
XX	cartilage differentiation; wound healing.
OS	Synthetic.
OS	Mus sp.
XX	
PN	W0200045831-A1.
PD	10-AUG-2000.
PF	02-FEB-2000; 2000WO-US02853.
PK	02-FEB-1999; 99US-0118276.
PA	(UYJE-) UNIV JEFFERSON THOMAS.
XX	
PI	San Antonio JD, Verrecchio A, Schick BP;
XX	WPI; 2000-543446/49.
DR	
XX	
PT	Novel synthetic peptides with high affinity for glycoaminoglycans and proteoglycans, useful for modulating heparin, promoting cell

PF attachment, modulating tumour metastasis and modulating wound healing -
XX
XX Claim 64; Page 58; 76pp; English.
XX
CC The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 90; DB 21; Length 15;
Best Local Similarity 100.0%; Pred. No. 9.4e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPARRARYQWVRCKP 15
DB 1 YPARRARYQWVRCKP 15
| | | | | | | | | | | | | | | | | |

RESULT 2
AAR42842
ID AAR42842 standard; Protein; 49 AA.
XX
AC AAR42842;
XX
DT 24-NOV-1993 (first entry)
XX
DE Mouse serglycin - exon 2 product.
XX
XX Haematopoietic cell secretory granule proteoglycan;
KW positive; negative; transcriptional regulatory element;
KW enhancer; eukaryotic promoter; constitutive suppressor;
KW TATA-box.
XX
OS Mus musculus.
XX
PN WO9313119-A.
XX
PD 08-JUL-1993.
XX
PF 23-DEC-1992; 92WO-US11194.
XX
PR 03-JAN-1992; 92US-0816289.
XX
PR 02-JUL-1992; 92US-0906871.
XX
PA (BGHM) BRIGHAM & WOMENS HOSPITAL.
XX
PI Abraham S, Stevens RL;
XX
DR WPI: 1993-227261/28.
XX
DR N-PSDB; AA050343.
XX
PT Transcription regulatory elements of ser-glycin gene - specific
PT for haematopoietic cells, also trans-acting transcriptional
PT binding factors
XX
PS Disclosure: Fig 5A; 112pp; English.
XX
XX Exons 1-3 of the mouse serglycin gene are given in AA044279 and
CC AA050343-44 and are interrupted by 8 kb and 4 kb sequences.
CC A negative transcription regulatory element (a constitutive
CC suppressor) between residues -250 and -190 of the 5' flanking

CC region of the mouse serglycin gene (see AA044279), a positive
CC (hematopoietic cell enhancer) regulatory element located between
CC residues -118 and -91, an equiv. of the TATA-box and a novel
CC eukaryotic promoter that utilises such equiv. are identified.
CC The regulatory elements, vectors and hosts provided with these
CC elements, are useful in the control of gene transcription of
CC heterologous genes in eukaryotic cells, esp. hematopoietic cells.
XX
SQ Sequence 49 AA;

Query Match 100.0%; Score 90; DB 14; Length 49;
Best Local Similarity 100.0%; Pred. No. 3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPARRARYQWVRCKP 15
DB 2 YPARRARYQWVRCKP 16
| | | | | | | | | | | | | | | | | |

RESULT 3
AAB08142
ID AAB08142 standard; peptide; 15 AA.
XX
AC AAB08142;
XX
DT 04-DEC-2000 (first entry)
XX
DE Serglycin peptide modulating activity of heparin, and other glycans.
XX
KW Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW cartilage differentiation; wound healing.
XX
XX Synthetic.
OS Homo sapiens.
XX
PN WO200045831-A1.
XX
PD 10-AUG-2000.
XX
PF 02-FEB-2000; 2000WO-US02853.
XX
PR 02-FEB-1999; 99US-0118276.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI San Antonio JD, Verrecchio A, Schlick BP;
XX
DR WPI: 2000-543446/49.
XX
XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
PS Claim 65; Page 58; 76pp; English.
XX
XX The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
XX
SQ Sequence 15 AA;

Db 25 YPTQRRARYQWVRCNP 39

RESULT 6
AAM81991

ID AAM81991 standard: Protein: 69 AA.

AC AAM81991;

DT 13-NOV-2001 (first entry)

DE Human haematological malignancy-related antigen #1689.

KW Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

OS Homo sapiens.

PN WO200164886-A2.

PD 07-SEP-2001.

PF 01-MAR-2001; 2001WO-US07272.

PR 01-MAR-2000; 2000US-0186126.

PR 17-MAR-2000; 2000US-0190479.

PR 27-APR-2000; 2000US-0200545.

PR 28-APR-2000; 2000US-0200303.

PR 28-APR-2000; 2000US-0200779.

PR 01-MAY-2000; 2000US-0200999.

PR 04-MAY-2000; 2000US-0202084.

PR 22-MAY-2000; 2000US-0206201.

PR 14-JUL-2000; 2000US-0218950.

PR 03-AUG-2000; 2000US-0222903.

PR 04-AUG-2000; 2000US-0223416.

PR 07-AUG-2000; 2000US-0223378.

XX (CORI-) CORIXA CORP.

PI Gaiger A, Algate PA, Mannion J;

DR WPI: 2001-514842/56.

XX Compositions and methods for the detection of hematological

PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular

PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -

PS Claim 1; Page 1183; 1252pp; English.

XX The present invention relates to compositions and methods for the

CC detection, diagnosis and therapy of haematological malignancies. The

CC present sequence is the protein sequence of a human haematological

CC malignancy related antigen. The methods of the present invention comprise

CC detecting the presence of haematological malignancy related antigen(s) in

CC a sample obtained from the patient (an increased level of the

CC polypeptide, compared to an unaffected individual, is indicative of an

CC increased risk). Haematological malignancies which can be treated using

CC the present invention are chronic lymphocytic leukaemia, lymphoma,

CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B

CC cell non-Hodgkin's lymphoma.

XX Sequence 69 AA:

SO Query Match 85.6%; Score 77; DB 22; Length 69;

Best Local Similarity 80.0%; Pred. No. 5e-05;

Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPARRRARYQWVRCNP 15

Db 25 YPTQRRARYQWVRCNP 39

RESULT 7
AAM80982

XX AAM80982 standard: Protein: 76 AA.

AC AAM80982;

DT 13-NOV-2001 (first entry)

DE Human haematological malignancy-related antigen #680.

KW Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

OS Homo sapiens.

PN WO200164886-A2.

PD 07-SEP-2001.

PF 01-MAR-2001; 2001WO-US07272.

PR 01-MAR-2000; 2000US-0186126.

PR 17-MAR-2000; 2000US-0190479.

PR 27-APR-2000; 2000US-0200545.

PR 28-APR-2000; 2000US-0200303.

PR 28-APR-2000; 2000US-0200779.

PR 01-MAY-2000; 2000US-0200999.

PR 04-MAY-2000; 2000US-0202084.

PR 22-MAY-2000; 2000US-0206201.

PR 14-JUL-2000; 2000US-0218950.

PR 03-AUG-2000; 2000US-0222903.

PR 04-AUG-2000; 2000US-0223416.

PR 07-AUG-2000; 2000US-0223378.

XX (CORI-) CORIXA CORP.

PI Gaiger A, Algate PA, Mannion J;

DR WPI: 2001-514842/56.

XX Compositions and methods for the detection of hematological

PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular

PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -

PS Claim 1; Page 785; 1252pp; English.

XX The present invention relates to compositions and methods for the

CC detection, diagnosis and therapy of haematological malignancies. The

CC present sequence is the protein sequence of a human haematological

CC malignancy related antigen. The methods of the present invention comprise

CC detecting the presence of haematological malignancy related antigen(s) in

CC a sample obtained from the patient (an increased level of the

CC polypeptide, compared to an unaffected individual, is indicative of an

CC increased risk). Haematological malignancies which can be treated using

CC the present invention are chronic lymphocytic leukaemia, lymphoma,

CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B

CC cell non-Hodgkin's lymphoma.

XX Sequence 76 AA:

SO Query Match 85.6%; Score 77; DB 22; Length 76;

Best Local Similarity 80.0%; Pred. No. 5.5e-05;

Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPARRRARYQWVRCNP 15

Db 32 YPTQRRARYQWVRCNP 46

RESULT 8
AAM81471

ID AAM81471 standard: Protein: 76 AA.

```

XX AC AAM81471;
XX DT 13-NOV-2001 (first entry)
XX DE Human haematological malignancy-related antigen #1169.
XX DE Human: cytostatic; vascular; gene therapy; vaccine; lymphoma;
XX DE haematological malignancy; antigen; chronic lymphocytic leukaemia;
XX DE follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX OS Homo sapiens.
XX PN WO200164886-A2.
XX PD 07-SEP-2001.
XX PF 01-MAR-2001; 2001WO-US07272.
XX PR 01-MAR-2000; 2000US-0186126.
XX PR 17-MAR-2000; 2000US-0190479.
XX PR 27-APR-2000; 2000US-0200545.
XX PR 28-APR-2000; 2000US-0200303.
XX PR 01-MAY-2000; 2000US-0200779.
XX PR 04-MAY-2000; 2000US-0202084.
XX PR 22-MAY-2000; 2000US-0206201.
XX PR 14-JUL-2000; 2000US-0218950.
XX PR 03-AUG-2000; 2000US-0222903.
XX PR 04-AUG-2000; 2000US-0223416.
XX PR 07-AUG-2000; 2000US-0223378.
XX PA (CORI-) CORIXA CORP.
XX PI Gaiger A, Algate PA, Mannion J;
XX DR WPI: 2001-514842/56.
XX PT Compositions and methods for the detection of hematological
XX PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
XX PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX PS Claim 1; Page 977; 1252pp; English.
XX CC The present invention relates to compositions and methods for the
XX CC detection, diagnosis and therapy of haematological malignancies. The
XX CC present sequence is the protein sequence of a human haematological
XX CC malignancy related antigen. The methods of the present invention comprise
XX CC detecting the presence of haematological malignancy related antigen(s) in
XX CC a sample obtained from the patient (an increased level of the
XX CC polypeptide, compared to an unaffected individual, is indicative of an
XX CC increased risk). Haematological malignancies which can be treated using
XX CC the present invention are chronic lymphocytic leukaemia, lymphoma,
XX CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
XX CC cell non-Hodgkin's lymphoma.
XX SQ Sequence 76 AA;
XX
XX Query Match 85.6%; Score 77; DB 22; Length 76;
XX Best Local Similarity 80.0%; Pred. No. 5.5e-05;
XX Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 YPARRARYQWRCNP 15
XX | : |||||
XX 32 YPTQRRARYQWRCNP 46
XX
XX RESULT 9
XX ID AAM81724
XX AC AAM81724;
XX DT 13-NOV-2001 (first entry)
XX DE Human haematological malignancy-related antigen #1165.

```

```

DT DT 13-NOV-2001 (first entry)
XX XX
XX DE Human haematological malignancy-related antigen #1422.
XX DE Human: cytostatic; vascular; gene therapy; vaccine; lymphoma;
XX DE haematological malignancy; antigen; chronic lymphocytic leukaemia;
XX DE follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX OS Homo sapiens.
XX PN WO200164886-A2.
XX PD 07-SEP-2001.
XX PF 01-MAR-2001; 2001WO-US07272.
XX PR 01-MAR-2000; 2000US-0186126.
XX PR 17-MAR-2000; 2000US-0190479.
XX PR 27-APR-2000; 2000US-0200545.
XX PR 28-APR-2000; 2000US-0200303.
XX PR 01-MAY-2000; 2000US-0200779.
XX PR 04-MAY-2000; 2000US-0202084.
XX PR 22-MAY-2000; 2000US-0206201.
XX PR 14-JUL-2000; 2000US-0218950.
XX PR 03-AUG-2000; 2000US-0222903.
XX PR 04-AUG-2000; 2000US-0223416.
XX PR 07-AUG-2000; 2000US-0223378.
XX PA (CORI-) CORIXA CORP.
XX PI Gaiger A, Algate PA, Mannion J;
XX DR WPI: 2001-514842/56.
XX PT Compositions and methods for the detection of hematological
XX PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
XX PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX PS Claim 1; Pages 1076-1077; 1252pp; English.
XX CC The present invention relates to compositions and methods for the
XX CC detection, diagnosis and therapy of haematological malignancies. The
XX CC present sequence is the protein sequence of a human haematological
XX CC malignancy related antigen. The methods of the present invention comprise
XX CC detecting the presence of haematological malignancy related antigen(s) in
XX CC a sample obtained from the patient (an increased level of the
XX CC polypeptide, compared to an unaffected individual, is indicative of an
XX CC increased risk). Haematological malignancies which can be treated using
XX CC the present invention are chronic lymphocytic leukaemia, lymphoma,
XX CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
XX CC cell non-Hodgkin's lymphoma.
XX SQ Sequence 76 AA;
XX
XX Query Match 85.6%; Score 77; DB 22; Length 76;
XX Best Local Similarity 80.0%; Pred. No. 5.5e-05;
XX Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 YPARRARYQWRCNP 15
XX | : |||||
XX 32 YPTQRRARYQWRCNP 46
XX
XX RESULT 10
XX ID AAM81954
XX AC AAM81954;
XX DT 13-NOV-2001 (first entry)
XX DE Human haematological malignancy-related antigen #1652.

```

```
XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
OS Homo sapiens.
XX WO200164886-A2.
PN
XX 07-SEP-2001.
PD
XX 01-MAR-2001; 2001WO-US07272.
PF
XX 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX (CORI-) CORIXA CORP.
PA
XX Gaiger A, Algate PA, Mannion J;
PI
XX WPI; 2001-514842/56.
DR
XX Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
XX Claim 1; Page 1168; 1252pp; English.
PS
XX
XX The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
XX Sequence 76 AA:
SQ
XX
XX Query Match 85.6%; Score 77; DB 22; Length 76;
XX Best Local Similarity 80.0%; Pred. No. 5.5e-05;
XX Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 YPARRARYQWVRKP 15
DB 32 YPTGRARYQWVRCP 46
II :|||||
RESULT 11
AAM80996
ID AAM80996 standard; Protein; 123 AA.
XX
XX AAM80996;
AC
XX 13-NOV-2001 (first entry)
DT
XX Human haematological malignancy-related antigen #694.
XX
XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
```

```
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
XX Homo sapiens.
OS
XX WO200164886-A2.
PN
XX 07-SEP-2001.
PD
XX 01-MAR-2001; 2001WO-US07272.
PF
XX 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX (CORI-) CORIXA CORP.
PA
XX Gaiger A, Algate PA, Mannion J;
PI
XX WPI; 2001-514842/56.
DR
XX Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
XX Claim 1; Page 791; 1252pp; English.
PS
XX
XX The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
XX Sequence 123 AA:
SQ
XX
XX Query Match 85.6%; Score 77; DB 22; Length 123;
XX Best Local Similarity 80.0%; Pred. No. 8.9e-05;
XX Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 YPARRARYQWVRKP 15
DB 29 YPTGRARYQWVRCP 43
II :|||||
RESULT 12
AAM81965
ID AAM81965 standard; Protein; 123 AA.
XX
XX AAM81965;
AC
XX 13-NOV-2001 (first entry)
DT
XX Human haematological malignancy-related antigen #1663.
XX
XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
XX Homo sapiens.
OS
```

```
XX XX WO200164886-A2.
XX PN
XX XX
XX PD 07-SEP-2001.
XX PF 01-MAR-2001; 2001WO-US07272.
XX XX
XX PR 01-MAR-2000; 2000US-0186126.
XX PR 17-MAR-2000; 2000US-0190479.
XX PR 27-APR-2000; 2000US-0200545.
XX PR 28-APR-2000; 2000US-0200303.
XX PR 28-APR-2000; 2000US-0200779.
XX PR 01-MAY-2000; 2000US-0200999.
XX PR 04-MAY-2000; 2000US-0202084.
XX PR 22-MAY-2000; 2000US-0206201.
XX PR 14-JUL-2000; 2000US-0218950.
XX PR 03-AUG-2000; 2000US-0222903.
XX PR 04-AUG-2000; 2000US-0223416.
XX PR 07-AUG-2000; 2000US-0223378.
XX XX
XX PA (CORI-) CORIXA CORP.
XX PI
XX PT Galger A, Algate PA, Mannion J;
XX DR WPI: 2001-514842/56.
XX XX
XX PT Compositions and methods for the detection of hematological
XX PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
XX PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX XX
XX PS Claim 1; Pages 1172-1173; 1252pp; English.
XX XX
XX CC The present invention relates to compositions and methods for the
XX CC detection, diagnosis and therapy of hematological malignancies. The
XX CC present sequence is the protein sequence of a human hematological
XX CC malignancy related antigen. The methods of the present invention comprise
XX CC detecting the presence of hematological malignancy related antigen(s) in
XX CC a sample obtained from the patient (an increased level of the
XX CC polypeptide, compared to an unaffected individual, is indicative of an
XX CC increased risk). Hematological malignancies which can be treated using
XX CC the present invention are chronic lymphocytic leukaemia, lymphoma,
XX CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
XX CC cell non-Hodgkin's lymphoma.
XX CC
XX SQ Sequence 123 AA;
XX XX
XX Query Match 85.6%; Score 77; DB 22; Length 123;
XX Best Local Similarity 80.0%; Pred. No. 8.9e-05;
XX Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 YPARRARYQWVRCNP 15
DB 29 YPTORARYQWVRCNP 43
DB

RESULT 13
AAM81008
ID AAM81008 standard; Protein: 143 AA.
XX AC
XX XX AAM81008;
XX DT 13-NOV-2001 (first entry)
XX DE Human haematological malignancy-related antigen #706.
XX KW Human; cytosolic; vascular; gene therapy; vaccine; lymphoma;
XX KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
XX KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX OS Homo sapiens.
XX XX
XX PN WO200164886-A2.
XX PF
```

```
PD 07-SEP-2001.
XX PF 01-MAR-2001; 2001WO-US07272.
XX XX
XX PR 01-MAR-2000; 2000US-0186126.
XX PR 17-MAR-2000; 2000US-0190479.
XX PR 27-APR-2000; 2000US-0200545.
XX PR 28-APR-2000; 2000US-0200303.
XX PR 28-APR-2000; 2000US-0200779.
XX PR 01-MAY-2000; 2000US-0200999.
XX PR 04-MAY-2000; 2000US-0202084.
XX PR 22-MAY-2000; 2000US-0206201.
XX PR 14-JUL-2000; 2000US-0218950.
XX PR 03-AUG-2000; 2000US-0222903.
XX PR 04-AUG-2000; 2000US-0223416.
XX PR 07-AUG-2000; 2000US-0223378.
XX XX
XX PA (CORI-) CORIXA CORP.
XX PI
XX PT Galger A, Algate PA, Mannion J;
XX DR WPI: 2001-514842/56.
XX XX
XX PT Compositions and methods for the detection of hematological
XX PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
XX PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX XX
XX PS Claim 1; Pages 796-797; 1252pp; English.
XX XX
XX CC The present invention relates to compositions and methods for the
XX CC detection, diagnosis and therapy of hematological malignancies. The
XX CC present sequence is the protein sequence of a human hematological
XX CC malignancy related antigen. The methods of the present invention comprise
XX CC detecting the presence of hematological malignancy related antigen(s) in
XX CC a sample obtained from the patient (an increased level of the
XX CC polypeptide, compared to an unaffected individual, is indicative of an
XX CC increased risk). Hematological malignancies which can be treated using
XX CC the present invention are chronic lymphocytic leukaemia, lymphoma,
XX CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
XX CC cell non-Hodgkin's lymphoma.
XX CC
XX SQ Sequence 143 AA;
XX XX
XX Query Match 85.6%; Score 77; DB 22; Length 143;
XX Best Local Similarity 80.0%; Pred. No. 0.0001;
XX Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 YPARRARYQWVRCNP 15
DB 29 YPTORARYQWVRCNP 43
DB

RESULT 14
AAM81975
ID AAM81975 standard; Protein: 143 AA.
XX AC
XX XX AAM81975;
XX DT 13-NOV-2001 (first entry)
XX DE Human haematological malignancy-related antigen #1673.
XX KW Human; cytosolic; vascular; gene therapy; vaccine; lymphoma;
XX KW haematological malignancy; antigen; chronic lymphocytic leukemia;
XX KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX OS Homo sapiens.
XX XX
XX PN WO200164886-A2.
XX PF 07-SEP-2001.
XX PD 01-MAR-2001; 2001WO-US07272.
XX PF
```

```
XX 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200959.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Algate PA, Mannion J;
XX
XX WPI; 2001-514842/56.
XX
XX Compositions and methods for the detection of hematological
XX malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
XX lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
XX Claim 1; Pages 1176-1177; 1252pp; English.
XX
XX
XX The present invention relates to compositions and methods for the
XX detection, diagnosis and therapy of hematological malignancies. The
XX present sequence is the protein sequence of a human haematological
XX malignancy related antigen. The methods of the present invention comprise
XX detecting the presence of hematological malignancy related antigen(s) in
XX a sample obtained from the patient (an increased level of the
XX polypeptide, compared to an unaffected individual, is indicative of an
XX increased risk). Haematological malignancies which can be treated using
XX the present invention are chronic lymphocytic leukaemia, lymphoma,
XX follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
XX cell non-Hodgkin's lymphoma.
XX
XX Sequence 143 AA:
SQ
XX
XX Query Match 85.6%; Score 77; DB 22; Length 143;
XX Best Local Similarity 80.0%; Pred. No. 0.0001;
XX Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 YPARRARYQWVRCKP 15
DB 29 YPTORARYQWVRCKP 43
XX
XX RESULT 15
XX AAM81000
XX ID AAM81000 standard; Protein; 148 AA.
XX
XX AAM81000;
XX
XX 13-NOV-2001 (first entry)
XX
XX Human haematological malignancy-related antigen #698.
XX
XX Human: cytostatic; vascular; gene therapy; vaccine; lymphoma;
XX haematological malignancy; antigen; chronic lymphocytic leukaemia;
XX follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
XX Homo sapiens.
XX
XX WO200164886-A2.
XX
XX 07-SEP-2001.
XX
XX 01-MAR-2001; 2001WO-US07272.
XX
XX 01-MAR-2000; 2000US-0186126.
XX
XX 17-MAR-2000; 2000US-0190479.
```

```
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200959.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Algate PA, Mannion J;
XX
XX WPI; 2001-514842/56.
XX
XX Compositions and methods for the detection of hematological
XX malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
XX lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
XX Claim 1; Page 793; 1252pp; English.
XX
XX
XX The present invention relates to compositions and methods for the
XX detection, diagnosis and therapy of hematological malignancies. The
XX present sequence is the protein sequence of a human haematological
XX malignancy related antigen. The methods of the present invention comprise
XX detecting the presence of hematological malignancy related antigen(s) in
XX a sample obtained from the patient (an increased level of the
XX polypeptide, compared to an unaffected individual, is indicative of an
XX increased risk). Haematological malignancies which can be treated using
XX the present invention are chronic lymphocytic leukaemia, lymphoma,
XX follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
XX cell non-Hodgkin's lymphoma.
XX
XX Sequence 148 AA:
SQ
XX
XX Query Match 85.6%; Score 77; DB 22; Length 148;
XX Best Local Similarity 80.0%; Pred. No. 0.0001;
XX Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 YPARRARYQWVRCKP 15
DB 32 YPTORARYQWVRCKP 46
XX
```

Search completed: April 8, 2003, 11:23:45
Job time : 64.3621 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:11:45 ; Search time 63.3621 Seconds
(without alignments)
31.545 Million cell updates/sec

Title: US-09-496-391-10
Perfect score: 92
Sequence: 1 YPTQRRARYQWVRCNP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	15	21	AAB08142
2	92	100.0	69	22	AAAB0105
3	92	100.0	69	22	AAAB1740
4	92	100.0	69	22	AAAB1991
5	92	100.0	76	22	AAAB0982
6	92	100.0	76	22	AAAB1471
7	92	100.0	76	22	AAAB1724
8	92	100.0	76	22	AAAB1954
9	92	100.0	123	22	AAAB0996
10	92	100.0	123	22	AAAB1965

11	92	100.0	143	22	AAAB1008	Human haematologic
12	92	100.0	143	22	AAAB1975	Human haematologic
13	92	100.0	148	22	AAAB1000	Human haematologic
14	92	100.0	156	22	AAAB1993	Human haematologic
15	92	100.0	158	11	AAAB05247	Peptide core of hu
16	92	100.0	158	11	AAAB05249	Proteoglycan pepi
17	92	100.0	158	14	AAAB3933	Serglycin - proteo
18	92	100.0	158	22	AAAB0992	Human haematologic
19	92	100.0	158	22	AAAB1961	Human haematologic
20	92	100.0	162	21	AAAB56784	Human prostrate can
21	86	93.5	158	21	AAAB03738	Human secretate pro
22	77	83.7	15	21	AAAB08141	Serglycin peptide
23	77	83.7	49	14	AAAB2842	Mouse serglycin -
24	75	81.5	164	22	AAAB30481	Novel human secret
25	73	79.3	16	22	AAAB99428	Vaccine related MH
26	68	73.9	201	22	AAAB1060	Novel human secret
27	59	64.1	125	22	AAAB0937	Human haematologic
28	59	64.1	125	22	AAAB1928	Human haematologic
29	45	48.9	122	23	ABG65138	Human albumin fusi
30	45	48.9	122	23	ABG65139	Human albumin fusi
31	45	48.9	122	23	ABB94022	Human secreted pro
32	45	48.9	122	23	ABB94046	Human secreted pro
33	45	48.9	176	21	AAAB44336	Arabidopsis thalia
34	45	48.9	183	21	AAAB17612	Arabidopsis thalia
35	45	48.9	183	21	AAAB38487	Arabidopsis thalia
36	45	48.9	206	22	AAAB47004	Plant D-like cycli
37	45	48.9	209	20	AAAB98179	Arabidopsis cyclin
38	45	48.9	209	21	AAAB17611	Arabidopsis thalia
39	45	48.9	209	21	AAAB38466	Arabidopsis thalia
40	44	47.8	54	21	AAAB26253	Corn cyclin-depend
41	44	47.8	292	23	AAAB1995	Herbidically activ
42	44	47.8	501	22	AAAB01153	Thermus rubens glu
43	43	46.7	85	22	ABG28206	Novel human diagno
44	43	46.7	95	22	ABG28205	Novel human diagno
45	43	46.7	99	22	ABG28204	Novel human diagno

ALIGNMENTS

RESULT 1	
AAAB08142	
AAAB08142 standard; peptide: 15 AA.	
XX	
AC	AAAB08142;
XX	
DT	04-DEC-2000 (first entry)
XX	
DE	Serglycin peptide modulating activity of heparin, and other glycans.
XX	
KW	Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW	cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW	cartilage differentiation; wound healing.
XX	
OS	Synthetic.
XX	
OS	Homo sapiens.
XX	
PM	WO200045831-A1.
XX	
PD	10-AUG-2000.
XX	
PF	02-FEB-2000; 2000WO-US02853.
XX	
PR	02-FEB-1999; 99US-0118276.
XX	
PA	(UYUE-) UNIV JEFFERSON THOMAS.
XX	
PI	San Antonio JD, Verrecchio A, Schick BP;
XX	
DR	WPI: 2000-543446/49.
XX	
PT	Novel synthetic peptides with high affinity for glycoaminoglycans and
PT	proteoglycans, useful for modulating heparin, promoting cell

PT		attachment, modulating tumour metastasis and modulating wound healing -
XX		
PS	Claim 65; Page 58; 76pp; English.	
CC	The present sequence represents a synthetic peptide which has a high	
CC	affinity for glycoaminoglycans and proteoglycans. The peptide is useful	
CC	in methods for modulating heparin or other glycoaminoglycans with	
CC	anticoagulant activity, promoting cell attachment or adhesion to	
CC	natural or synthetic surfaces (especially vein grafts), modulating	
CC	tumour cell metastasis, modulating cartilage differentiation, targeting	
CC	drugs to epithelial cell surfaces (or to other cells expressing	
CC	proteoglycans), modulating enzymes that act on glycoaminoglycan	
CC	substrates, affinity purification of bioactive sequences of a	
CC	glycoaminoglycan, modifying endothelial cell pro-coagulant or	
CC	anti-coagulant functions mediated through glycoaminoglycans, and	
CC	modulating wound healing. The peptide may also be used for blocking	
CC	tissue uptake of heparin or other glycoaminoglycans in a mammal to	
CC	Increase heparin half-life in circulation.	
XX		
SQ	Sequence 15 AA;	
	Query Match 100.0%; Score 92; DB 21; Length 15;	
	Best Local Similarity 100.0%; Pred. No. 2.1e-08;	
	Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
OY	1 YPTORARYQWVRGNP 15 1 YPTORARYQWVRGNP 15	
Db		
	RESULT 2	
	AAM81015	
AC	AAM81015 standard; Protein; 69 AA.	
XX	AAM81015;	
XX		
DT	13-NOV-2001 (first entry)	
XX		
DE	Human haematological malignancy-related antigen #713.	
XX		
KW	Human; cytosolic; vascular; gene therapy; vaccine; lymphoma;	
KW	haematological malignancy; antigen; chronic lymphocytic leukemia;	
KW	follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.	
XX		
OS	Homo sapiens.	
XX		
FN	WO200164886-A2.	
XX		
PD	07-SEP-2001.	
XX		
FE	01-MAR-2001; 2001WO-US07272.	
XX		
PR	01-MAR-2000; 2000US-0186126.	
PR	17-MAR-2000; 2000US-0190479.	
PR	27-APR-2000; 2000US-0200545.	
PR	28-APR-2000; 2000US-0200303.	
PR	28-APR-2000; 2000US-0200779.	
PR	01-MAY-2000; 2000US-0200999.	
PR	04-MAY-2000; 2000US-0202084.	
PR	22-MAY-2000; 2000US-0206201.	
PR	14-JUL-2000; 2000US-0218950.	
PR	03-AUG-2000; 2000US-0222903.	
PR	04-AUG-2000; 2000US-0223416.	
PR	07-AUG-2000; 2000US-0223378.	
XX		
PA	(CORI-) CORIXA CORP.	
XX		
PI	Gaiger A, Algate PA, Mannion J;	
XX		
DR	WPI: 2001-514842/56.	
XX		
PT	Compositions and methods for the detection of hematological	
XX	malignancies e.g. chronic lymphocytic leukemia, lymphoma, follicular	

```

PT Lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
PS Claim 1; Page 799; 1252pp; English.
XX
CC The present invention relates to compositions and methods for the
XX detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
SQ Sequence 69 AA;
XX
Query Match 100.0%; Score 92; DB 22; Length 69;
Best Local Similarity 100.0%; Pred. NO. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY 1 YPTQRRARYQWVRCNP 15
| | | | | | | | | | | | | | |
DB 25 YPTQRRARYQWVRCNP 39
RESULT 3
AAM81740
ID AAM81740 standard; Protein; 69 AA.
AC AAM81740;
XX
DT 13-NOV-2001 (first entry)
DE Human haematological malignancy-related antigen #1438.
XX
KW Human; cytosstatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; Chronic lymphocytic Leukemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
OS Homo sapiens.
XX
PN WO200164886-A2.
PD
PP 07-SEP-2001.
XX
PF 01-MAR-2001; 2001WO-US07272.
XX
PR 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 12-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
PA (CORI-) CORIXA CORP.
XX
PI Galger A, Algate PA, Mannion J;
XX
WP1: 2001-514842/56.
XX
PT Compositions and methods for the detection of hematological
XX malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
XX lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
PS Claim 1; Page 1082; 1252pp; English.

```


XX The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.

SQ Sequence 69 AA:

Query Match 100.0%; Score 92; DB 22; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPTQRRARYQWVRNCP 15
|||||
DB 25 YPTQRRARYQWVRNCP 39

RESULT 4
AAM81991
ID AAM81991 standard; Protein; 69 AA.
AC AAM81991;
XX
XX 13-NOV-2001 (first entry)
XX
XX Human haematological malignancy-related antigen #1689.
DE
XX
XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
XX Homo sapiens.
OS
XX
XX WO200164886-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 01-MAR-2001; 2001WO-US07272.
PF
XX
XX 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Gaiger A, Algate PA, Mannion J;
PI
XX
XX WPI: 2001-514842/56.
DR
XX
XX Compositions and methods for the detection of haematological
PT malignancies, e.g. chronic lymphocytic leukaemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
XX Claim 1; Page 1183; 1252pp; English.
PS
XX
XX The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The

CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.

SQ Sequence 69 AA:

Query Match 100.0%; Score 92; DB 22; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPTQRRARYQWVRNCP 15
|||||
DB 25 YPTQRRARYQWVRNCP 39

RESULT 5
AAM80982
ID AAM80982 standard; Protein; 76 AA.
AC AAM80982;
XX
XX 13-NOV-2001 (first entry)
XX
XX Human haematological malignancy-related antigen #680.
DE
XX
XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
XX Homo sapiens.
OS
XX
XX WO200164886-A2.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 01-MAR-2001; 2001WO-US07272.
PF
XX
XX 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Gaiger A, Algate PA, Mannion J;
PI
XX
XX WPI: 2001-514842/56.
DR
XX
XX Compositions and methods for the detection of haematological
PT malignancies, e.g. chronic lymphocytic leukaemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
XX Claim 1; Page 785; 1252pp; English.
PS
XX
XX The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in

CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.

XX Sequence 76 AA;

Query Match 100.0%; Score 92; DB 22; Length 76;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTQARYQWRCP 15
| | | | | | | | | | | | | | | | | |

Db 32 YPTQARYQWRCP 46

RESULT 6

AAM81471

ID AAM81471 standard; Protein; 76 AA.

AC AAM81471;

DT 13-NOV-2001 (first entry)

DE Human haematological malignancy-related antigen #1169.

XX Human: cytostatic; vascular; gene therapy; vaccine; lymphoma;

KW haematological malignancy; antigen; chronic lymphocytic leukaemia;

KM follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

XX Homo sapiens.

PN WO200164886-A2.

PD 07-SEP-2001.

PE 01-MAR-2001; 2001MO-US07272.

PR 01-MAR-2000; 2000US-0186126.

PR 17-MAR-2000; 2000US-0190479.

PR 27-APR-2000; 2000US-0200545.

PR 28-APR-2000; 2000US-0200303.

PR 28-APR-2000; 2000US-0200779.

PR 01-MAY-2000; 2000US-0200999.

PR 04-MAY-2000; 2000US-0202084.

PR 22-MAY-2000; 2000US-0206201.

PR 14-JUL-2000; 2000US-0218950.

PR 03-AUG-2000; 2000US-0222903.

PR 04-AUG-2000; 2000US-0223416.

PR 07-AUG-2000; 2000US-0223378.

PA (CORI-) CORIXA CORP.

PI Gaiger A, Algate PA, Mannion J;

DR WPI; 2001-514842/56.

PT Compositions and methods for the detection of hematological

PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular

PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -

XX Claim 1; Page 977; 1252pp; English.

XX The present invention relates to compositions and methods for the
XX detection, diagnosis and therapy of haematological malignancies. The
XX present sequence is the protein sequence of a human haematological

CC malignancy related antigen. The methods of the present invention comprise

CC detecting the presence of haematological malignancy related antigen(s) in

CC a sample obtained from the patient (an increased level of the

CC polypeptide, compared to an unaffected individual, is indicative of an

CC increased risk). Haematological malignancies which can be treated using

CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.

XX Sequence 76 AA;

Query Match 100.0%; Score 92; DB 22; Length 76;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTQARYQWRCP 15
| | | | | | | | | | | | | | | | | |

Db 32 YPTQARYQWRCP 46

RESULT 7

AAM81724

ID AAM81724 standard; Protein; 76 AA.

AC AAM81724;

DT 13-NOV-2001 (first entry)

DE Human haematological malignancy-related antigen #1422.

XX Human: cytostatic; vascular; gene therapy; vaccine; lymphoma;

KW haematological malignancy; antigen; chronic lymphocytic leukaemia;

KM follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

XX Homo sapiens.

PN WO200164886-A2.

PD 07-SEP-2001.

PE 01-MAR-2001; 2001MO-US07272.

PR 01-MAR-2000; 2000US-0186126.

PR 17-MAR-2000; 2000US-0190479.

PR 27-APR-2000; 2000US-0200545.

PR 28-APR-2000; 2000US-0200303.

PR 28-APR-2000; 2000US-0200779.

PR 01-MAY-2000; 2000US-0200999.

PR 04-MAY-2000; 2000US-0202084.

PR 22-MAY-2000; 2000US-0206201.

PR 14-JUL-2000; 2000US-0218950.

PR 03-AUG-2000; 2000US-0222903.

PR 04-AUG-2000; 2000US-0223416.

PR 07-AUG-2000; 2000US-0223378.

PA (CORI-) CORIXA CORP.

PI Gaiger A, Algate PA, Mannion J;

DR WPI; 2001-514842/56.

PT Compositions and methods for the detection of hematological

PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular

PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -

XX Claim 1; Pages 1076-1077; 1252pp; English.

XX The present invention relates to compositions and methods for the
XX detection, diagnosis and therapy of haematological malignancies. The
XX present sequence is the protein sequence of a human haematological

CC malignancy related antigen. The methods of the present invention comprise

CC detecting the presence of haematological malignancy related antigen(s) in

CC a sample obtained from the patient (an increased level of the

CC polypeptide, compared to an unaffected individual, is indicative of an

CC increased risk). Haematological malignancies which can be treated using

CC the present invention are chronic lymphocytic leukaemia, lymphoma,

CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B

CC cell non-Hodgkin's lymphoma.

XX Sequence 76 AA; 100.0%; Score 92; DB 22; Length 76;
SQ Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTORARYQWRCNP 15
|||||
Db 32 YPTORARYQWRCNP 46

RESULT 8
ID AAM81954 standard; Protein: 76 AA.
XX AAM81954;
AC AAM81954;
XX 13-NOV-2001 (first entry)
DT
XX Human haematological malignancy-related antigen #1652.
DE
XX Human; cytosolic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX Homo sapiens.
OS
XX WO200164886-A2.
PN
XX 07-SEP-2001.
PD
XX 01-MAR-2001; 2001WO-US07272.
PF
XX 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX (CORI-) CORIXA CORP.
PA
XX Gaiger A, Algate PA, Mannion J;
PI
XX WPI; 2001-514842/56.
DR
XX Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
PS
XX Claim 1; Page 1168; 1252pp; English.
PS
XX The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
XX Sequence 76 AA;

Query Match 100.0%; Score 92; DB 22; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTORARYQWRCNP 15
|||||
Db 32 YPTORARYQWRCNP 46

RESULT 9
ID AAM80996 standard; Protein: 123 AA.
XX AAM80996;
AC AAM80996;
XX 13-NOV-2001 (first entry)
DT
XX Human haematological malignancy-related antigen #694.
DE
XX Human; cytosolic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX Homo sapiens.
OS
XX WO200164886-A2.
PN
XX 07-SEP-2001.
PD
XX 01-MAR-2001; 2001WO-US07272.
PF
XX 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX (CORI-) CORIXA CORP.
PA
XX Gaiger A, Algate PA, Mannion J;
PI
XX WPI; 2001-514842/56.
DR
XX Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
PS
XX Claim 1; Page 791; 1252pp; English.
PS
XX The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
XX Sequence 123 AA;

Query Match 100.0%; Score 92; DB 22; Length 123;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTORARYQWVRNCP 15
DB 29 YPTORARYQWVRNCP 43
RESULT 10
AAM81965
ID AAM81965 standard; Protein; 123 AA.
AC AAM81965;
XX
XX
DT 13-NOV-2001 (first entry)
DE Human haematological malignancy-related antigen #1663.
XX
XX
KW Human: cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
OS Homo sapiens.
XX
PN WO200164886-A2.
XX
PD 07-SEP-2001.
XX
PF 01-MAR-2001; 2001WO-US07272.
XX
PR 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0206201.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Algate PA, Mannion J;
XX
DR WPI; 2001-514842/56.
XX
XX
PT Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
XX
PS Claim 1; Pages 1172-1173; 1252pp; English.
XX
XX
CC The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of hematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
XX
SQ Sequence 123 AA;
Query Match 100.0%; Score 92; DB 22; Length 123;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 29 YPTORARYQWVRNCP 43
RESULT 11
AAM81008
ID AAM81008 standard; Protein; 143 AA.
AC AAM81008;
XX
XX
DT 13-NOV-2001 (first entry)
DE Human haematological malignancy-related antigen #706.
XX
XX
KW Human: cytostatic; vascular; gene therapy; vaccine; lymphoma;
KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
OS Homo sapiens.
XX
PN WO200164886-A2.
XX
PD 07-SEP-2001.
XX
PF 01-MAR-2001; 2001WO-US07272.
XX
PR 01-MAR-2000; 2000US-0186126.
PR 17-MAR-2000; 2000US-0190479.
PR 27-APR-2000; 2000US-0200545.
PR 28-APR-2000; 2000US-0200303.
PR 28-APR-2000; 2000US-0200779.
PR 01-MAY-2000; 2000US-0200999.
PR 04-MAY-2000; 2000US-0202084.
PR 22-MAY-2000; 2000US-0218950.
PR 14-JUL-2000; 2000US-0218950.
PR 03-AUG-2000; 2000US-0222903.
PR 04-AUG-2000; 2000US-0223416.
PR 07-AUG-2000; 2000US-0223378.
XX
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Algate PA, Mannion J;
XX
DR WPI; 2001-514842/56.
XX
XX
PT Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
XX
PS Claim 1; Pages 796-797; 1252pp; English.
XX
XX
CC The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
XX
SQ Sequence 143 AA;
Query Match 100.0%; Score 92; DB 22; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
RESULT 12
AA081975
ID AA081975 standard; Protein; 143 AA.
XX
AC AA081975;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human haematological malignancy-related antigen #1673.
XX
KW Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KM haematological malignancy; antigen; chronic lymphocytic leukaemia;
XX follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
OS Homo sapiens.
XX
PN WO200164886-A2.
XX
PD 07-SEP-2001.
XX
PF 01-MAR-2001; 2001WO-US07272.
XX
PR 01-MAR-2000; 2000US-0186126.
XX
PR 17-MAR-2000; 2000US-0190479.
XX
PR 27-APR-2000; 2000US-0200545.
XX
PR 28-APR-2000; 2000US-0200303.
XX
PR 28-APR-2000; 2000US-0200779.
XX
PR 01-MAY-2000; 2000US-0200999.
XX
PR 04-MAY-2000; 2000US-0202084.
XX
PR 22-MAY-2000; 2000US-0206201.
XX
PR 14-JUL-2000; 2000US-0218950.
XX
PR 03-AUG-2000; 2000US-0222903.
XX
PR 04-AUG-2000; 2000US-0223416.
XX
PR 07-AUG-2000; 2000US-0223378.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Algate PA, Mannion J;
XX
DR WPI; 2001-514842/56.
XX
PT Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
PS Claim 1; Pages 1176-1177; 1252pp; English.
XX
CC The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
SQ Sequence 143 AA;
XX
Query Match 100.0%; Score 92; DB 22; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YPTQRRARYQWVRCNP 15
DB 29 YPTQRRARYQWVRCNP 43
RESULT 13
AA081000
ID AA081000 standard; Protein; 148 AA.
```

```
XX
AC AA081000;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human haematological malignancy-related antigen #698.
XX
KW Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
KM haematological malignancy; antigen; chronic lymphocytic leukaemia;
XX follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.
XX
OS Homo sapiens.
XX
PN WO200164886-A2.
XX
PD 07-SEP-2001.
XX
PF 01-MAR-2001; 2001WO-US07272.
XX
PR 01-MAR-2000; 2000US-0186126.
XX
PR 17-MAR-2000; 2000US-0190479.
XX
PR 27-APR-2000; 2000US-0200545.
XX
PR 28-APR-2000; 2000US-0200303.
XX
PR 28-APR-2000; 2000US-0200779.
XX
PR 01-MAY-2000; 2000US-0200999.
XX
PR 04-MAY-2000; 2000US-0202084.
XX
PR 22-MAY-2000; 2000US-0206201.
XX
PR 14-JUL-2000; 2000US-0218950.
XX
PR 03-AUG-2000; 2000US-0222903.
XX
PR 04-AUG-2000; 2000US-0223416.
XX
PR 07-AUG-2000; 2000US-0223378.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Algate PA, Mannion J;
XX
DR WPI; 2001-514842/56.
XX
PT Compositions and methods for the detection of hematological
PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
XX
PS Claim 1; Page 793; 1252pp; English.
XX
CC The present invention relates to compositions and methods for the
CC detection, diagnosis and therapy of haematological malignancies. The
CC present sequence is the protein sequence of a human haematological
CC malignancy related antigen. The methods of the present invention comprise
CC detecting the presence of haematological malignancy related antigen(s) in
CC a sample obtained from the patient (an increased level of the
CC polypeptide, compared to an unaffected individual, is indicative of an
CC increased risk). Haematological malignancies which can be treated using
CC the present invention are chronic lymphocytic leukaemia, lymphoma,
CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
CC cell non-Hodgkin's lymphoma.
XX
SQ Sequence 148 AA;
XX
Query Match 100.0%; Score 92; DB 22; Length 148;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 YPTQRRARYQWVRCNP 15
DB 32 YPTQRRARYQWVRCNP 46
RESULT 14
AA081993
ID AA081993 standard; Protein; 156 AA.
XX
AC AA081993;
XX
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: April 8, 2003, 11:17:38 ; Search time 53.2759 Seconds
(without alignments)
58.013 Million cell updates/sec

Title: US-09-496-391-10
Perfect score: 92
Sequence: 1 YPTQRRARYQWVRCNP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriap: *
17: SP archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	51.1	415	16	Q92P21 rhizobium m
2	45	48.9	177	17	Q8TNA7
3	45	48.9	183	16	Q8R8X7
4	45	48.9	209	10	Q9SCR2
5	44	47.8	231	10	Q9SV79
6	44	47.8	292	10	Q80781
7	44	47.8	755	10	Q9FGM2
8	44	47.8	1383	2	Q9K5L9
9	43	46.7	119	2	Q51446
10	43	46.7	120	16	Q9HXC3
11	43	46.7	354	5	Q9XV21
12	43	46.7	421	5	Q46123
13	43	46.7	438	5	Q46124
14	43	46.7	505	16	Q92Y56
15	43	46.7	782	5	Q18654
16	43	46.7	822	5	Q96106

17	43	46.7	951	5	Q9VDH2	O9vdh2 drosophila
18	42	45.7	233	10	Q9SV73	Q9sv73 arabidopsis
19	42	45.7	351	10	Q92RH6	Q92rh6 rhizobium m
20	42	45.7	383	5	Q9GUM2	Q9gum2 caenorhabdi
21	42	45.7	416	16	Q8XB29	Q8xb29 escherichia
22	42	45.7	484	17	Q8TWJ9	Q8twj9 methanopyru
23	41.5	45.1	678	16	Q8ZOP1	Q8zop1 anabaena sp
24	41	44.6	131	5	Q9TW98	Q9tw98 pinctada fu
25	41	44.6	131	5	Q9TW92	Q9tw92 pinctada fu
26	41	44.6	421	16	Q981A5	Q981a5 rhizobium 1
27	41	44.6	428	3	Q05814	Q05814 saccharomye
28	41	44.6	446	5	Q01754	Q01754 caenorhabdi
29	41	44.6	476	10	Q04927	Q04927 allium tube
30	41	44.6	1474	5	Q9N9K4	Q9n9k4 leishmania
31	41	44.6	1708	17	Q26769	Q26769 methanobact
32	41	44.6	2703	10	Q9MA24	Q9ma24 arabidopsis
33	41	44.6	2810	10	Q9FKS4	Q9fks4 arabidopsis
34	40.5	44.0	92	6	Q8WMB8	Q8wmb8 lagenerhyn
35	40.5	44.0	956	5	Q93LM5	Q93lm5 nostoc punc
36	40.5	44.0	956	5	Q95V61	Q95v61 drosophila
37	40.5	44.0	1937	2	Q8RJY2	Q8rjy2 stigmatella
38	40.5	44.0	3546	2	Q9P830	Q9p830 micromonos
39	40	43.5	91	5	Q9N3P7	Q9n3p7 caenorhabdi
40	40	43.5	125	16	Q8ZOK7	Q8zok7 anabaena sp
41	40	43.5	125	16	Q9RCV3	Q9rcv3 streptomyce
42	40	43.5	129	5	Q97048	Q97048 pinctada fu
43	40	43.5	129	5	Q9UHA3	Q9uah3 pinctada fu
44	40	43.5	132	12	Q8QRX6	Q8qrx6 chimpanzee
45	40	43.5	233	2	Q93P78	Q93p78 microscilla

ALIGNMENTS

RESULT 1
ID Q92P21 PRELIMINARY; PRT; 415 AA.
AC Q92P21;
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE putative transport transmembrane protein.
GN R01579 OR SMC01212.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothé G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Gofrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger F.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
RL EMBL: AL591787; CAC46158.1;
DR InterPro: IPR000566; Lipocln_cytfABP.
DR InterPro: IPR003662; sub.transporter.
DR Pfam: PF00083; sugar.tr.1.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 415 AA; 43992 MW; 8C22BFBEB9314AC2 CRC64;
Query Match 51.1%; Score 47; DB 16; Length 415;
Best Local Similarity 50.0%; Pred. No. 9.6;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 2 PTORRARYQWVRCNP 15

Db 193 PANRREHMRNP 206

RESULT 2

08TNA7 PRELIMINARY; PRT; 177 AA.
AC 08TNA7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Predicted protein.
GN MA386.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A/ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Altschuld D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearlano K., Johnson R.,
RA Linton L., McKean P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grane D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuelner H.C., Kirycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL: AE010929; AAM05772.1; -;
KW Complete proteome.
SQ SEQUENCE 177 AA; 20878 MW; CAD2FFP290DD9F51 CRC64;
QY 1 YPTORARYOWRCN 14
Db 115 YDSEKNOYKWLKCS 128
Query Match 48.9%; Score 45; DB 17; Length 177;
Best Local Similarity 35.7%; Pred. No. 8.7;
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
QY 1 YPTORARYOWRCN 14
Db 115 YDSEKNOYKWLKCS 128
RESULT 3
08RBY7 PRELIMINARY; PRT; 183 AA.
AC 08RBY7;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein TTE0679.
GN TTE0679.
OS Thermobacterium tengcongensis.
OC Bacteria; Firmicutes; Bacilli/Clostridium group; Clostridia;
OC Thermobacteriales; Thermobacteriaceae; Thermobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
DR EMBL: AE013036; AAM23943.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 183 AA; 21048 MW; FD21BC165B30606B CRC64;

Query Match 48.9%; Score 45; DB 16; Length 183;
Best Local Similarity 70.0%; Pred. No. 9;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 TORARYOWVR 12
Db 80 TRRSRYWVR 89

RESULT 4

09SCR2 PRELIMINARY; PRT; 209 AA.
AC 09SCR2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE Hypothetical 24.0 kDa protein (Cdc2a-interacting protein).
GN T3A5.10 OR ICK2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloeker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Lui H., Wang H., Delong C., Fowke L.C., Crosby W.L., Robert P.R.;
RT "The Arabidopsis cdc2a-interacting protein ICK2 is structurally
RT related to ICK1 and is a potent inhibitor of cyclin-dependent kinase
RT activity in vitro."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132979; CAB62432.1; -;
DR EMBL: AJ251851; CAB76424.1; -;
KW Hypothetical protein.
SQ SEQUENCE 209 AA; 24036 MW; 0E5EEF9ED256B53 CRC64;
QY 2 PTQARYOWVRNP 15
Db 196 PLGGREYWKLPN 209
Query Match 48.9%; Score 45; DB 10; Length 209;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 2 PTQARYOWVRNP 15
Db 196 PLGGREYWKLPN 209
RESULT 5
09SV79 PRELIMINARY; PRT; 231 AA.
AC 09SV79;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 26.0 kDa protein.
GN F25G13.3 OR AT4G12900.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Scheller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-13 FROM N.A.
 RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Mewes H.W.,
 RA Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Robben J., Gymnopoulos B., Volckaert G., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL079349; CAB53090.1; -
 DR EMBL; AL161535; CAB78332.1; -
 DR InterPro: IPR004911; GILT.
 DR InterPro: IPR000834; Zn_carboxypept.
 DR Pfam: PF03227; GILT. 1.
 DR PROSITE: PS00133; CARBOXYPEPT_2N_2; UNKNOWN_1.
 DR Hypothetical protein.
 KW SEQUENCE 231 AA; 26025 MW; 734109A78E942295 CRC64;

Query Match 47.8%; Score 44; DB 10; Length 231;
 Best Local Similarity 46.2%; Pred. No. 17;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 YPTQARYQWVRC 13
 DB 110 WPNQRLHYKFTRC 122
 ID 080781 PRELIMINARY; PRT; 292 AA.
 AC 080781;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Putative non-LTR retroelement reverse transcriptase.
 GN AT2G34320.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Taiton L.D., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Niernman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana";
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AC004481; AAC27408.1; -
 KW RNA-directed DNA polymerase.
 SQ SEQUENCE 292 AA; 34221 MW; 1B7D3866C4C9C7 CRC64;

Query Match 47.8%; Score 44; DB 10; Length 292;
 Best Local Similarity 85.7%; Pred. No. 21;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 YQWVRCN 14
 DB 142 YQWVRCN 148
 ID 09K519 PRELIMINARY; PRT; 755 AA.
 AC 09K519;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE DnaJ protein-like.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB025622; BAB08418.1; -
 DR HSP; P25685; 1HDJ.
 DR InterPro: IPR001623; DnaJ_N.
 DR InterPro: IPR003095; Hsp_DnaJ.
 DR Pfam: PF00226; DnaJ. 1.
 DR PRINTS: PR00625; DnaJPROTEIN.
 DR SMART: SM00271; DnaJ. 1.
 DR PROSITE: PS00636; DnaJ_1; UNKNOWN_1.
 DR PROSITE: PS50076; DnaJ_2; 1.
 SQ SEQUENCE 755 AA; 85237 MW; 47C58BDF62560FE CRC64;

Query Match 47.8%; Score 44; DB 10; Length 755;
 Best Local Similarity 50.0%; Pred. No. 54;
 Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 YPTQARYQWVRCN 14
 DB 196 YPTQESSTFWTCN 209
 ID 09K519 PRELIMINARY; PRT; 1383 AA.
 AC 09K519;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Peptide synthetase.
 GN ADP.
 OS Anabaena sp. 90.
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
 OX NCBI_TaxID=46234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=90;
 RX MEDLINE=20392447; PubMed=10931313;
 RA Rouhainen L., Paulin L., Suomalainen S., Hyttinen H., Bulkema W.,
 RA Haselkorn R., Sivonen K.;
 RT "Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in Anabaena strain 90."
 RL MOL. MICROBIOL. 37:156-167(2000).
 DR EMBL; AJ269505; CAC01606.1; -
 DR HSP; P14687; 1AMU.
 DR InterPro: IPR000873; AMP-bind.

DR InterPro: IPR001242; Condensatn.
 DR InterPro: IPR000010; Cystatin.
 DR InterPro: IPR003880; Ppanine_attach.
 DR InterPro: IPR001031; Thioesterase.
 DR Pfam: PF00501; AMP-binding: 1.
 DR Pfam: PF00668; Condensation: 1.
 DR Pfam: PF00550; pp-binding: 1.
 DR Pfam: PF00975; Thioesterase: 1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00075; ACP_DOMAIN: 1.
 DR PROSITE: PS00455; AMP_BINDING: 1.
 DR PROSITE: PS00287; CYSTATIN; UNKNOWN_1.
 KM Phosphopentetheine.
 SQ SEQUENCE 1383 AA; 155394 MW; C9E9C2D68F43AEF CRC64;

Query Match 47.8%; Score 44; DB 2; Length 1383;
 Best Local Similarity 60.0%; Pred. No. 1e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPTORARYOW 10
 |||:|:|
 Db 586 YPTERLERFMW 595

RESULT 9
 O51446 PRELIMINARY; PRT; 119 AA.
 AC O51446;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE Exoenzyme S (Exos) (Fragment).
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=388;
 RX MEDLINE=95173091; Pubmed=7868588;
 RA Yahr T.L., Hovey A.K., Kulich S.M., Frank D.W.;
 RT "Transcriptional analysis of the Pseudomonas aeruginosa exoenzyme S structural gene";
 RL J. Bacteriol. 177:1169-1178(1995).
 DR EMBL: L27629; AAA66489.1; -.
 FT NON_TER 119 119
 SO SEQUENCE 119 AA; 12479 MW; 858259770FD01151 CRC64;

Query Match 46.7%; Score 43; DB 2; Length 119;
 Best Local Similarity 53.8%; Pred. No. 12;
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 TORARYOWRCNP 15
 ||:|:|
 Db 103 TOTSSRWVSCNP 115

RESULT 10
 O9HXG3 PRELIMINARY; PRT; 120 AA.
 AC O9HXG3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE Hypothetical protein PA3843.
 GN PA3843.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAO1;

RX MEDLINE=20437337; Pubmed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL: AE004801; AAG07230.1; -.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 120 AA; 12580 MW; 1E9A73A862B6462A CRC64;

Query Match 46.7%; Score 43; DB 16; Length 120;
 Best Local Similarity 53.8%; Pred. No. 13;
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 TORARYOWRCNP 15
 ||:|:|
 Db 103 TOTSSRWVSCNP 115

RESULT 11
 O9XV21 PRELIMINARY; PRT; 354 AA.
 AC O9XV21;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE F46B3.9 protein.
 GN F46B3.9.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderiinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ainscough R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; Pubmed=9851916;
 RA none;
 RT "Investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL: Z81540; CAB04398.1; -.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR003645; FOLN.
 DR SMART: SM00001; EGF_like; 1.
 DR SMART: SM00274; FOLN; 6.
 SO SEQUENCE 354 AA; 37229 MW; 5769BF095EBE03F9 CRC64;

Query Match 46.7%; Score 43; DB 5; Length 354;
 Best Local Similarity 71.4%; Pred. No. 37;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 9 QWVRCNP 15
 ||:|:|
 Db 90 QWIKCNP 96

RESULT 12
 O46123 PRELIMINARY; PRT; 421 AA.
 AC O46123;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE GBR-2A protein precursor.
 GN GBR-2.
 OS Haemonchus contortus (Barber pole worm).

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Strongylida;
 OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
 OX NCBI_TaxID=6289;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20017520; PubMed=10551358;
 RA Jagannathan S., Laughton D.L., Critten C.L., Skinner T.M.,
 RA Horoszk L., Wolstenholme A.D.;
 RT "ligand-gated chloride channel subunits encoded by the Haemonchus
 RT contortus and Ascaris suum orthologues of the Caenorhabditis
 RT elegans.";
 RL MOL. Biochem. Parasitol. 103:129-140(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Jagannathan S.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 DR EMBL: Y14233; CAA74622.2; -;
 DR InterPro: IPR000188; GABAA_receptor.
 DR InterPro: IPR001175; Neur_channel.
 DR Pfam: PF02932; Neur_chan_LBD; 1.
 DR PRINTS: PR00252; NRIONCHANNEL.
 DR TIGRFAMs: TIGR00860; LIC; 1.
 DR PROSITE: PS00236; NEUROTR_ION_CHANNEL; 1.
 KW Glycoprotein; Ionic channel; Postsynaptic membrane; Signal;
 FT SIGNAL. 1 26 POTENTIAL.
 SQ SEQUENCE 421 AA; 48557 MW; D13AE87D9D4E8348 CRC64;
 Query Match 46.7%; Score 43; DB 5; Length 421;
 Best Local Similarity 46.7%; Pred. No. 44;
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 OY 1 YPTQARYQWRCNP 15
 Db 201 YTTQDIKYEKQNP 215
 RESULT 13
 O46124
 ID O46124 PRELIMINARY; PRT; 438 AA.
 AC O46124;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE GBR-28 protein precursor.
 GN GBR-2.
 OS Haemonchus contortus (Barber pole worm).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Strongylida;
 OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
 OX NCBI_TaxID=6289;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20017520; PubMed=10551358;
 RA Jagannathan S., Laughton D.L., Critten C.L., Skinner T.M.,
 RA Horoszk L., Wolstenholme A.D.;
 RT "ligand-gated chloride channel subunits encoded by the Haemonchus
 RT contortus and Ascaris suum orthologues of the Caenorhabditis
 RT elegans.";
 RL MOL. Biochem. Parasitol. 103:129-140(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 DR EMBL: Y14234; CAA74623.1; -;
 DR InterPro: IPR000188; GABAA_receptor.
 DR InterPro: IPR001175; Neur_channel.
 DR Pfam: PF02932; Neur_chan_LBD; 1.
 DR PRINTS: PR00252; NRIONCHANNEL.
 DR TIGRFAMs: TIGR00860; LIC; 1.
 DR PROSITE: PS00236; NEUROTR_ION_CHANNEL; 1.
 KW Glycoprotein; Ionic channel; Postsynaptic membrane; Signal;

KW Transmembrane.
 FT SIGNAL. 1 26 POTENTIAL.
 SQ SEQUENCE 438 AA; 50690 MW; D679BD6BCF67FE1D CRC64;
 Query Match 46.7%; Score 43; DB 5; Length 438;
 Best Local Similarity 46.7%; Pred. No. 46;
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
 OY 1 YPTQARYQWRCNP 15
 Db 201 YTTQDIKYEKQNP 215
 RESULT 14
 O92Y56
 ID O92Y56 PRELIMINARY; PRT; 505 AA.
 AC O92Y56;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative reverse transcriptase.
 GN RA1032 OR SMA1875.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSymb (megaplasmid 1).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396509; PubMed=11481432;
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gonzy J.,
 RA Gujal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
 RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 RT "Nucleotide sequence and predicted functions of the entire
 RT Sinorhizobium meliloti pSymb megaplasmid.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 DR EMBL: AE007289; AAK65690.1; -;
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF00078; rvt; 1.
 DR PRINTS: PR01561; EDG8RECEPTOR.
 KW RNA-directed DNA polymerase; Plasmid; Hypothetical protein;
 KW Complete proteome.
 SQ SEQUENCE 505 AA; 57987 MW; 69D684807B96273A CRC64;
 Query Match 46.7%; Score 43; DB 16; Length 505;
 Best Local Similarity 58.3%; Pred. No. 53;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 OY 4 QRAYQWRCNP 15
 Db 16 QKIXQWSKNP 27
 RESULT 15
 Q18654
 ID Q18654 PRELIMINARY; PRT; 782 AA.
 AC Q18654; Q9T2R5;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical 87.4 kDa protein (Host cell factor 1).
 GN C46A5.9 OR HCF-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
 OC Rhabdilitidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:12:02 ; Search time 11.6379 Seconds
(Without alignments)
53.458 Million cell updates/sec

Title: US-09-496-391-10
Perfect score: 92
Sequence: 1 YPTQRRARYQWVRCNP 15

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	92	100.0	158	1	PGSG_HUMAN
2	77	83.7	152	1	PGSG_MOUSE
3	77	83.7	179	1	PGSG_RAT
4	42	45.7	313	1	NUDC_MYCTU
5	42	45.7	416	1	YIDR_ECOLI
6	41.5	45.1	203	1	R15B_YEAST
7	40	43.5	512	1	ACCD_YEAST
8	40	43.5	522	1	ACCD_TOBAC
9	39	42.4	150	1	YH8W_YEAST
10	39	42.4	398	1	C142_MYCTU
11	39	42.4	504	1	YC62_CHYU
12	39	42.4	539	1	Y469_HUMAN
13	39	42.4	566	1	YBAE_ECOLI
14	39	42.4	2205	1	POLN_RUBVT
15	38	41.3	72	1	VG35_BPML5
16	38	41.3	672	1	LOX5_RAT
17	38	41.3	673	1	LOX5_HUMAN
18	38	41.3	725	1	CYPC_MYCLE
19	38	41.3	780	1	PRTP_HSV1F
20	38	41.3	785	1	PRTP_HSV1I
21	38	41.3	785	1	PRTP_HSV1A
22	38	41.3	1403	1	YCN1_YEAST
23	38	41.3	3770	1	ACVS_YEMENI
24	37.5	40.8	229	1	YCS9_MYCLE
25	37	40.2	252	1	POQC_ACTIC
26	37	40.2	275	1	CHER_VIBAN
27	37	40.2	275	1	CHER_VIBPA
28	37	40.2	275	1	CHRI_VIBCH
29	37	40.2	333	1	PDXA_CAUCR
30	37	40.2	335	1	NU2M_LUMRE
31	37	40.2	386	1	Y155_AQDAE
32	37	40.2	398	1	CYH1_HUMAN
33	37	40.2	433	1	ELT2_CAEBL

34	37	40.2	466	1	GAC2_RAT	P18508 rattus norv
35	37	40.2	467	1	GAC2_HUMAN	P18507 homo sapien
36	37	40.2	474	1	GAC2_CHICK	P21548 gallus gall
37	37	40.2	474	1	GAC2_MOUSE	P22723 mus musculu
38	37	40.2	475	1	GAC2_BOVIN	P22300 bos taurus
39	37	40.2	581	1	PRLR_BOVIN	Q28172 bos taurus
40	37	40.2	581	1	PRLR_SHEEP	O46561 ovis aries
41	37	40.2	664	1	EMAP_LYTA	O9Y1C1 lytechinus
42	37	40.2	669	1	GSH1_SCHPO	O09768 schizosacch
43	37	40.2	686	1	EMAP_STRPU	Q26613 strongyloc
44	37	40.2	738	1	112R_MOUSE	O60837 mus musculu
45	37	40.2	743	1	2267_HUMAN	Q14586 homo sapien

ALIGNMENTS

```

RESULT 1
PGSG_HUMAN
ID PGSG_HUMAN STANDARD: PRT; 158 AA.
AC P10124:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Secretory granule proteoglycan core protein precursor (Platelet
DE proteoglycan core protein) (P.PG) (Hematopoietic proteoglycan core
DE protein) (Serglycin).
CN PRGI OR PG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed=2798108;
RX MEDLINE=90016819; PubMed=2798108;
RA Stellericht C.M., Saunders G.F.;
RT "Nucleotide sequence of a cDNA encoding a hemopoietic proteoglycan
RT core protein.";
RL Nucleic Acids Res. 17:7523-7523(1989).
RN [2]
RP SEQUENCE FROM N.A. PubMed=2835370;
RX MEDLINE=88213411; PubMed=2835370;
RA Stevens R.L., Avraham S., Gartner M.C., Bruns G.A.P., Austen K.F.,
RA Weis J.H.;
RT "Isolation and characterization of a cDNA that encodes the peptide
RT core of the secretory granule proteoglycan of human promyelocytic
RT leukemia HL-60 cells.";
RL J. Biol. Chem. 263:7287-7291(1988).
RN [3]
RP SEQUENCE FROM N.A. PubMed=2180935;
RX MEDLINE=90202841; PubMed=2180935;
RA Nicodemus C.F., Avraham S., Austen K.F., Purdy S., Jablonski J.,
RA Stevens R.L.;
RT "Characterization of the human gene that encodes the peptide core of
RT secretory granule proteoglycans in promyelocytic leukemia HL-60 cells
RT and analysis of the translated product.";
RL J. Biol. Chem. 265:5889-5896(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX TISSUE=Lung;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 28-158, AND SEQUENCE OF 34-158 FROM N.A.
RX MEDLINE=88296856; PubMed=3402609;
RA Alliel P.M., Perin J.-P., Maillet P., Bonnet F., Rosa J.-P.,
RA Jolles P.;
RT "Complete amino acid sequence of a human platelet proteoglycan.";
RL FEBS Lett. 236:123-126(1988).
RN [6]
RP SEQUENCE OF 28-93.
RX MEDLINE=89104992; PubMed=3214420;
RA Perin J.-P., Bonnet F., Maillet P., Jolles P.;

```



```

ID PGSG_RAT STANDARD: PRT: 179 AA.
AC P04917;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Secretary granule proteoglycan core protein precursor (Chondroitin
DE sulfate proteoglycan core protein) (Proteoglycan 10K core protein)
DE (pc19 core protein) (Cytolytic granule proteoglycan core protein).
DE PR01.
CS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=86213412; Pubmed=3366780;
RA Avraham S., Stevens R.L., Gartner M.C., Austen K.F., Lalley P.A.,
RA Weis J.H.;
RT "Isolation of a cDNA that encodes the peptide core of the secretory
RT granule proteoglycan of rat basophilic leukemia-1 cells and
RT assessment of its homology to the human analogue."
RL J. Biol. Chem. 263:7292-7296(1988).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=yoik sac carcinoma;
RC MEDLINE=86304425; Pubmed=2427521;
RX Bourdon M.A., Shiga M., Ruoslahti E.;
RA "Identification from cDNA of the precursor form of a chondroitin
RT sulfate proteoglycan core protein."
RL J. Biol. Chem. 261:12534-12537(1986).
RN [3]
RN SEQUENCE FROM N.A.
RP MEDLINE=90279692; Pubmed=2352541;
RX Glorda R., Chambers W.H., Dahl C.A., Trucco M.;
RA "Isolation and characterization of a cDNA that encodes the core
RT protein of the cytolitic granule proteoglycan in rat natural killer
RT cells."
RL Nat. Immun. Cell Growth Regul. 9:91-102(1990).
RN [4]
RN SEQUENCE OF 75-179 FROM N.A.
RP MEDLINE=85140287; Pubmed=3919394;
RX Bourdon M.A., Oldberg A., Pletschbacher M.D., Ruoslahti E.;
RA "Molecular cloning and sequence analysis of a chondroitin sulfate
RT proteoglycan cDNA."
RL Proc. Natl. Acad. Sci. U.S.A. 82:1321-1325(1985).
CC -1- SIMILARITY: TO CORRESPONDING PROTEINS IN HUMAN AND MOUSE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02934; AAA42171.1; -
CC DR EMBL; J03224; AAA41837.1; -
CC DR PIR; A25644; GZPT0
CC DR PIR; A28059; A28059.
CC DR PIR; A60653; A60653.
CC KW Glycoprotein; Proteoglycan; Repeat; Signal.
CC FT SIGNAL 1 26 POTENTIAL.
CC FT PROPEP 27 75 ACTIVATION PEPTIDE.
CC FT CHAIN 76 179 SECRETORY GRANULE PROTEOGLYCAN CORE
CC FT PROTEIN 24 X 2 AA TANDEM REPEATS OF S-G.
CC FT DISULFID 39 48 POTENTIAL.
CC SQ SEQUENCE 179 AA; 18577 MW; D2E2AE7D3AAA0D6 CRC64;
Query March 83.7%; Score 77; DB 1; length 179;
Best local similarity 80.0%; Pred. No. 8.3e-06;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

07 YPTORARYOWNRCPN 15
|| :|||||
Db 27 YPARRARYOWNRCKP 41

RESULT 4
NUDC_MYCTU
ID NUDC_MYCTU STANDARD: PRI: 313 AA.
AC 053345;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NAH pyrophosphatase (EC 3.6.1.-).
SD NUDC OR RV3199C OR MT3293 OR MTFV014.43C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteriidae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S., Hornsbly T., Jagels K., Krogan A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RT Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh:
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.P., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains." Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RL -I CARAYACTIC ACTIVITY: NADH + H(2)O = AMP + NMNH.
CC -I COFACTOR: REQUIRES DIVALENT IONS: MANGANESE OR MAGNESIUM (BY SIMILARITY).
CC -I SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -I SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY. NUDC SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).
CC -----
DR EMBL: AL021646; CAAL6664.1; -
DR EMBL: AE007142; AAAK47636.1; ALT_INTIT.
DR TIGR: MT3293; -
DR TubercuList; RV3199C; -
DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam: PF00293; NUDIX_1.
DR PRINTS: PR00502; NUDIXFAMILY.
DR PROSITE: PS00893; NUDIX_1.
DR HydroLase; NAD; Magnesium; Manganese; Complete proteome.
FT DOMAIN 203 224 NUDIX BOX.
SO SEQUENCE 313 AA; 33826 MW; 52E33E252FEF803 CRC64;

Query Match 45.7%; Score 42; DB 1; Length 313;
Best Local Similarity 57.1%; Read NO. 8.7;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 PTOBARYQWVRCNP 15
11 11 11 11
DB 147 PTKPARAGMSRVNP 160

RESULT 5

ID YIDR_ECOLI STANDARD; PRT; 416 AA.
AC P31455; P76734;
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yidR.
GN YIDR OR B3689.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. IIR, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
genome: organizational symmetry around the origin of replication.";
RL Genomes 16:551-561(1993).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L10328; AAA62041.1; -
DR EMBL: AE000446; AAC76712.1; -
DR EcoGene: EG11713; yidR.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 416 AA; 46318 MW; 47E36360CC99AB7 CRC64;

Query Match 45.7%; Score 42; DB 1; Length 416;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 WVRNCP 15
111111
DB 302 WVRNCP 307

RESULT 6
R15B_YEAST
ID R15B_YEAST STANDARD; PRT; 203 AA.
AC P54780;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L15-B (YL10) (L13) (Rpl15R) (Yp18).
GN RPL15B OR RPL10B OR YL10B OR YMR121C OR YMR564.03C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Lye G., Churcher C.M., Barrell B.G., Rajandream M.A.;
RT Submitted (May-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=83048950; PubMed=6814480;
RA Otake E., Higo K.-I., Osawa S.;

RT "Isolation of seventeen proteins and amino-terminal amino acid
RT sequences of eight proteins from cytoplasmic ribosomes of yeast.";
RL Biochemistry 21:4545-4550(1982).
RN [3]

RP SEQUENCE OF 1-8.

RX MEDLINE=92184799; PubMed=1544921;
RA Takakura H., Tsunashima S., Miyagi M., Warner J.R.;
RT "NH2-terminal acetylation of ribosomal proteins of Saccharomyces
RL cerevisiae.";
RN J. Biol. Chem. 267:5442-5445(1992).

CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L15 IN YEAST.

CC -1- SIMILARITY: BELONGS TO THE L15E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z49273; CAA89270.1; -
DR SGD: S0004728; RPL15B.
DR InterPro: IPR000439; Ribosomal_L15e.
DR Pfam: PF00827; Ribosomal_L15e; 1.
DR PROSITE: PS01194; RIBOSOMAL_L15E; 1.
KW Ribosomal protein; Multigene family.
FT INIT. MET 0
FT CONFIRM 27
SQ SEQUENCE 203 AA; 24291 MW; AE472EB562931E2C CRC64;

Query Match 45.1%; Score 41.5; DB 1; Length 203;
Best Local Similarity 70.0%; Pred. No. 6.8;
Matches 7; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

OY 6 ARQWVRCNP 15
111 1 111
DB 145 ARYNNMI-CNP 153

RESULT 7
ACCD_TOBAC
ID ACCD_TOBAC STANDARD; PRT; 512 AA.
AC P12219;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
DE (EC 6.4.1.2) (ACCase beta chain).
GN ACCD OR YCF11 OR ZFPA.
OS Nicotiana tabacum (Common tobacco).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bright yellow 4;
RA Shinzaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N.,
RA Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,
RA Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B.-Y., Sugita M.,
RA Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
RA Tohdo N., Shimada H., Sugiura M.;
RT "The complete nucleotide sequence of the tobacco chloroplast genome:
its gene organization and expression."

EMBO J. 5:2043-2049(1986).
CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate
CC + malonyl-CoA.
CC -1- SIMILARITY: BELONGS TO THE ACCD / PCOB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: 200044; CAA7362.1; -
DR PIR: A05196; A05196.
DR InterPro: IPR000438; ACOACC_transfB.
DR InterPro: IPR000022; Carboxyl_trans.
DR Pfam: PF01039; Carboxyl_trans; 1.
DR PRINTS: PR01070; ACCCTFRASEB.
DR TIGRfams: TIGR00515; accd; 1.
KW Fatty acid biosynthesis; ligase; zinc-finger; Chloroplast.
FT ZN_FING 247 269 C4-TYPE (POTENTIAL).
SQ SEQUENCE 512 AA; 58472 MW; DFGC312E622868 CRC64;

Query Match 43.5%; Score 40; DB 1; Length 512;
Best Local Similarity 54.5%; Pred. No. 31;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 TORARYQWYRC 13
||:|:|:|:|
DB 237 TQRYHLMWQC 247

RESULT 8

ACCD_SPIOL STANDARD; PRT; 522 AA.
AC 09M3L7;

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
DE (EC 6.4.1.2) (ACCase beta chain).
GN ACCD.

OS Spinacia oleracea (Spinach).
OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_Taxid=3562;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=cv. Geant d'hiver, and cv. Monatol;
RA MEDLINE=21187424; PubMed=11292076;
RA Schmitz-Linneweber C., Maier R.M., Alcaraz J.-P., Cottet A.,
RA Hermann R.G., Mache R.;

RT "The plastid chromosome of spinach (Spinacia oleracea): complete
RT nucleotide sequence and gene organization";
RL Plant Mol. Biol. 45:307-315(2001).
CC -i- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.

CC -i- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate
CC + malonyl-CoA.
CC -i- SIMILARITY: BELONGS TO THE ACCD / PCOB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ400848; CAB88738.1; -
DR InterPro: IPR000438; ACOACC_transfB.
DR InterPro: IPR000022; Carboxyl_trans.
DR Pfam: PF01039; Carboxyl_trans; 1.
DR PRINTS: PR01070; ACCCTFRASEB.
DR TIGRfams: TIGR00515; accd; 1.

KW Fatty acid biosynthesis; ligase; zinc-finger; Chloroplast.
FT ZN_FING 258 280 C4-TYPE (POTENTIAL).
SQ SEQUENCE 522 AA; 58907 MW; 84A337E7B6A3BF87 CRC64;

Query Match 43.5%; Score 40; DB 1; Length 522;
Best Local Similarity 54.5%; Pred. No. 31;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 TORARYQWYRC 13
||:|:|:|:|
DB 248 TQRYHLMWQC 258

RESULT 9

YMBW_YEAST STANDARD; PRT; 150 AA.
ID YMBW_YEAST

AC 003579;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 17.2 kDa protein in PRC1-AD64 intergenic region.
GN YMR298W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_Taxid=4932;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972.
RA Badcock K., Churcher C., Barrell B.G., Raeburn M.A.;
RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X80836; CAA56807.1; -
DR SGG: S0004913; YMR298W.
KW Hypothetical protein.
SQ SEQUENCE 150 AA; 17207 MW; 75DBA35225C3065C CRC64;

Query Match 42.4%; Score 39; DB 1; Length 150;
Best Local Similarity 45.5%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 5 RARYQWYRCNP 15
||:|:|:|:|
DB 45 RINTWPHCTP 55

RESULT 10

C142_MYCTU STANDARD; PRT; 398 AA.
ID C142_MYCTU

AC 053563;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative cytochrome P450 142 (EC 1.14.-.-).
GN CYP142 OR RV3518C OR MW3619 OR MW023.25C.

OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacteriinae; Mycobacteriaceae; Mycobacterium.

OX NCBI_Taxid=1773;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=96295987; PubMed=9634220;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

```

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornbly T., Jagels K., Kiroh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Salston J.E., Taylor K., Whitehead S., Barrell B.G.:
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=CDCC 1551 / Oshkosh:
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwin M.L., Helt D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.:
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL022022: CAA17755.1; -
DR EMBL: AB007165: AAK47979.1; ALT_INIT.
DR HSSP: P33006: ICPT.
DR TIGR: MT3619; -
DR TubercuList: RV3518C; -
DR InterPro: IPR001128: Cytochrome_P450.
DR Pfam: PF00067: P450.1.
DR PRINTS: PR00385: P450.
DR PROSITE: PS00086: CYTOCHROME_P450; 1.
DR Hypothetical protein: Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT BINDING 340 340 HEME (BY SIMILARITY).
FT CONFLICT 1 18 MTEAPVDLADGNFYASR -> GMSPNRSAPPSSRRARP
FT (IN REF. 2).
SQ SEQUENCE 398 AA: 44398 MW: BCF93C23ECB5767F CRC64;

Query Match 42.4%; Score 39; DB 1; Length 398;
Best Local Similarity 60.0%; Pred. No. 35;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 RARYQWVRCN 14
DB 21 RAAVYRWVRAN 30

RESULT 11
YC62_CHLVU STANDARD; PRT; 504 AA.
AC 020163;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Hypothetical 60.0 kDa protein ycf62 (ORF504).
GN YCF62.
OS Chlorella vulgaris.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Chlorella.
OX NCBI_TaxID=3077;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IAM C-27 / Tamiya;
MEDLINE=97303241; PubMed=9159184;
RX

```

```

RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
RA Tsudzuki J., Nishikawa K., Tsudzuki T., Suzuki Y., Hamada A., Ohba T.,
RA Inamura A., Yoshinaga K., Sugita M.:
RT "Complete nucleotide sequence of the chloroplast genome from the
RT green alga Chlorella vulgaris: the existence of genes possibly
RT involved in chloroplast division."
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
CC -1- SIMILARITY: BELONGS TO THE UPF0072 (MSU/YCF62) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB001684: BAA57935.1; -
DR Hypothetical protein: Chloroplast.
SQ SEQUENCE 504 AA: 59972 MW: E2A4FD28192C6AF0 CRC64;

Query Match 42.4%; Score 39; DB 1; Length 504;
Best Local Similarity 35.7%; Pred. No. 44;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 YPTQRAYQWVRCN 14
DB 426 FPTKKKKRWFYCN 439

RESULT 12
Y469_HUMAN STANDARD; PRT; 539 AA.
AC 090J94; 075057;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0469.
GN KIAA0469.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,
RA Nakajima D., Nomura N., Ohara O.:
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain."
RL DNA Res. 4:345-349(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Wray P.:
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB007938: BAA32314.1; -
DR EMBL: AL031447: CAB42892.1; -
DR InterPro: IPR000210: BTB_POZ.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00651: BTB; 1.
DR Pfam: PF01344; Kelch; 5.
RX

```

```

DR SMART; SM00225; BTB; 1.
DR PROSITE; PS50097; BTB; 1.
KW Hypothetical protein; Repeat.
FT DOMAIN 35 103
FT REPEAT 287 335 BTB
FT REPEAT 336 382 KEICH 1.
FT REPEAT 384 422 KEICH 2.
FT REPEAT 423 470 KEICH 3.
FT REPEAT 472 512 KEICH 4.
FT REPEAT 514 534 KEICH 5.
FT CONFLICT 539 AA; 60435 MW; DDB88D184C077945 CRC64;
SQ SEQUENCE 539 AA; 60435 MW; DDB88D184C077945 CRC64;

Query Match 42.4%; Score 39; DB 1; Length 539;
Best Local Similarity 44.4%; Pred. No. 47;
Matches 8; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

OY 2 PTOARATQ---WVRCNP 15
DB 193 PKEEAAYQIALRWVRADP 210

RESULT 13
YBAE_ECOLI STANDARD; PRT; 566 AA.
AC P46890; P71158;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ybae.
GN YBAE OR B0445.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Siao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Hatada E., Omori H., Qiao Y., Tsuji M., Fukuda R.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-358 FROM N.A.
RC STRAIN=K12 / MC4100;
RA Patzer S.I., Hantke K.;
RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
CC -i SIMILARITY: TO E.COLI YABN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE000151; AAC73548.1; -
CC EMBL; U82664; AAB40201.1; -

```

```

DR EMBL; D82943; BA011649.1; -
DR EMBL; Z54355; CA091182.1; -
DR EcoGene; EG13217; ybae.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 566 AA; 64991 MW; F25DC077252C51D1 CRC64;

Query Match 42.4%; Score 39; DB 1; Length 566;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 9 QWVRCNP 15
DB 474 QWLRCDP 480

RESULT 14
POLN_RUBVT STANDARD; PRT; 2205 AA.
AC P13889;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nonstructural polyprotein [Contains: Nonstructural protein NSP1;
DE Nonstructural protein NSP2; Nonstructural protein NSP3; Nonstructural
DE protein NSP4]
OS Rubella virus (strain Therien).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Rubivirus.
OX NCBI_TaxID=11045;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90281585; PubMed=2353453;
RA Dominguez G., Wang C.Y., Frey T.K.;
RT "Sequence of the genome RNA of rubella virus: evidence for genetic
RT rearrangement during togavirus evolution."
RL Virology 177:225-258(1990).
RN [2]
RP SEQUENCE OF 1737-2205 FROM N.A.
RC MEDLINE=88226020; PubMed=2836271;
RA Frey T.K., Marr L.D.;
RT "Sequence of the region coding for virion proteins C and E2 and the
RT carboxy terminus of the nonstructural proteins of rubella virus:
RL comparison with alphaviruses."
RL Gene 62:85-99(1988).
CC -i PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M15240; AAA8528.1; -
CC PIR; A35320; MNWVRN.
DR MEROPS; C27.001; -
DR InterPro; IPR002589; A1PP.
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR SMART; SM00506; A1PP; 1.
KW Polyprotein; Nonstructural protein.
FT CHAIN 1 ? NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 2 ? NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 3 ? NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 4 ? NONSTRUCTURAL PROTEIN NSP4.
FT CHAIN 5 ? NONSTRUCTURAL PROTEIN NSP5.
SQ SEQUENCE 2205 AA; 240220 MW; 15A95F55E34C0B03 CRC64;

```

Query Match 42.4%; Score 39; DB 1; Length 2205;
 Best Local Similarity 50.0%; Pred. No. 2e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 YPTQRRARYQW 10
 :||:|:|:|
 Db 1462 WPTFRSRHTW 1471

RESULT 15
 VG35_BPML5

ID VG35_BPML5 STANDARD; PRT; 72 AA.

AC 005245;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last annotation update)

DE Gene 35 protein (GP35).

GN 35.

OS Mycobacteriophage L5.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;

OC L5-like viruses.

OX NCBI_TaxID=31757;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93211282; PubMed=8459766;

RA Hatfull G.F., Sarkis G.J.;

RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
 a phage system for mycobacterial genetics.";

RL Mol. Microbiol. 7:395-405(1993).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: Z18946; CAA79411.1; -

DR PIR: S30980; S30980.

SQ SEQUENCE 72 AA; 8264 MW; 782B42ED99C5D0E0 CRC64;

Query Match 41.3%; Score 38; DB 1; Length 72;
 Best Local Similarity 46.7%; Pred. No. 8.8;
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 YPTQRRARYQWRCNP 15
 |||:|:|:|
 Db 5 YPPOQDRPQWYAGSP 19

Search completed: April 8, 2003, 11:24:42
 Job time : 13.6379 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:18:53 ; Search time 23.7931 Seconds
(without alignments)
60.606 Million cell updates/sec

Title: US-09-496-391-10

Perfect score: 92

Sequence: 1 YP09ARYQWRCNP 15

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_73:*
2: pir1:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	158	2 A28058	secretory granule
2	77	83.7	152	2 T00791	secretory granule 17K c
3	77	83.7	179	1 G2R70	secretory granule
4	45	48.9	209	2 T46140	hypothetical prote
5	44	47.8	91	2 T06502	hypothetical prote
6	44	47.8	231	2 H85138	hypothetical prote
7	44	47.8	292	2 T02320	hypothetical prote
8	43	46.7	120	2 D83166	hypothetical prote
9	43	46.7	354	2 T22274	hypothetical prote
10	43	46.7	505	2 H95390	hypothetical prote
11	43	46.7	717	2 T29816	probable reverse t
12	43	46.7	782	2 T43277	hypothetical prote
13	42	45.7	233	2 T10198	hypothetical prote
14	42	45.7	313	2 E70951	hypothetical prote
15	42	45.7	416	2 B65171	hypothetical prote
16	42	45.7	416	2 E91207	hypothetical prote
17	42	45.7	416	2 G86053	hypothetical prote
18	41.5	45.1	204	2 S54490	hypothetical prote
19	41.5	45.1	678	2 A11812	ribosomal protein
20	41	44.6	448	2 T15188	DNA mismatch repair
21	41	44.6	1708	2 T59189	hypothetical prote
22	41	44.6	125	2 AD1817	protoporphyrin IX
23	40	43.5	343	2 T29547	hypothetical prote
24	40	43.5	512	2 A05196	hypothetical prote
25	40	43.5	925	2 T02811	hypothetical prote
26	39	42.4	150	2 S47459	DNA excision/repair
27	39	42.4	196	2 T09968	probable membrane
28	39	42.4	204	2 T01447	cyclin-dependent k
29	39	42.4	293	2 T06027	xyloglucan endo-1,

30	39	42.4	398	1 H70807	cytochrome P450 Rv
31	39	42.4	408	2 B75635	phosphoenolpyruvat
32	39	42.4	504	2 T07287	hypothetical prote
33	39	42.4	566	2 C90691	hypothetical prote
34	39	42.4	566	2 G85541	hypothetical prote
35	39	42.4	566	2 E64774	hypothetical prote
36	39	42.4	575	2 T08964	probable membrane
37	39	42.4	719	2 A87292	hypothetical prote
38	39	42.4	896	2 AE1785	primosomal protein
39	39	42.4	896	2 AE1409	the two components
40	39	42.4	2205	1 MNWVRN	the two components
41	38	41.3	72	2 S30980	nonstructural poly
42	38	41.3	105	2 T18036	gene 35 protein -
43	38	41.3	111	2 T14306	hypothetical prote
44	38	41.3	265	2 P85014	glycine-rich prote
45	38	41.3	266	2 A48349	hypothetical prote
					U28 protein - ate

ALIGNMENTS

RESULT 1
A28058
secretory granule proteoglycan core protein precursor - human
N:Alternate names: HL-60 cell proteoglycan peptide core; platelet proteoglycan
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 13-Aug-1999
C:Accession: A35183; A28058; S09610; S01126; S01819
R:Nicodemus, C.F.; Avraham, S.; Austen, K.F.; Purdy, S.; Jablonski, J.; Stevens, R.L.
J. Biol. Chem. 265, 5889-5896, 1990
A:Title: Characterization of the human gene that encodes the peptide core of secretor
A:Reference number: A35183; MUID:90202841; PMID:2180935
A:Accession: A35183
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-158 <NIC>
A:Cross-references: GB:J05824
R:Stevens, R.L.; Avraham, S.; Gartner, M.C.; Bruns, G.A.P.; Austen, K.F.; Weis, J.H.
J. Biol. Chem. 263, 7287-7291, 1988
A:Title: Isolation and characterization of a cDNA that encodes the peptide core of th
A:Reference number: A28058; MUID:88213411; PMID:2835370
A:Accession: A28058
A:Molecule type: mRNA
A:Residues: 1-158 <STE>
A:Cross-references: GB:J03223; NID:g190419; PIDN:AAA60179.1; PID:g190420
R:Stellrecht, C.M.; Sanders, G.F.
Nucleic Acids Res. 17, 7523, 1989
A:Title: Nucleotide sequence of a cDNA encoding a hemopoietic proteoglycan core prote
A:Reference number: S09610; MUID:90016819; PMID:2798108
A:Accession: S09610
A:Molecule type: mRNA
A:Residues: 1-158 <ST2>
A:Cross-references: EMBL:X17042; NID:g32432; PIDN:CAA34900.1; PID:g32433
R:Alliel, P.M.; Perlin, J.P.; Maillet, P.; Bonnet, F.; Rosa, J.P.; Jolles, P.
FEBS Lett. 236, 123-126, 1988
A:Title: Complete amino acid sequence of a human platelet proteoglycan.
A:Reference number: S01126; MUID:88296856; PMID:3402609
A:Accession: S01126
A:Molecule type: mRNA
A:Residues: 28-138, 'S', 140-158 <ALD>
A:Cross-references: EMBL:X12765
A>Note: part of this sequence was confirmed by protein sequencing
A>Note: 139-Asn was also found
R:Perlin, J.P.; Bonnet, F.; Maillet, P.; Jolles, P.
Biochem. J. 255, 1007-1013, 1988
A:Title: Characterization and N-terminal sequence of human platelet proteoglycan.
A:Reference number: S01819; MUID:89104992; PMID:3214420
A:Accession: S01819
A:Molecule type: protein
A:Residues: 28-48, 'X', 50-93 <PER>
C:Genetics:
A:Gene: GDB:PRG1; PRG
A:Cross-references: GDB:120312; OMIM:177040

A:Map position: 10q22.1-10q22.1
C:Superfamily: proteoglycan 10K core protein
C:Keywords: chondroitin sulfate proteoglycan
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-158/Product: secretory granule proteoglycan core protein #status predicted <MAT>

Query Match 100.0%; Score 92; DB 2; Length 158;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTORARYQWVRCNP 15
DB 28 YPTORARYQWVRCNP 42

RESULT 2

proteoglycan 17K core protein precursor - mouse
N:Alternate names: secretory granule proteoglycan core protein; serglycin
C:Species: Mus musculus (house mouse)
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 13-Aug-1999
C:Accession: J00791; A34399; S06127; A32779
R:Avraham, S.; Huang, R.; Aveskog, M.; Petersson, I.; Kjellen, L.; Hellman, L.
Gene 93, 235-240, 1990
A:Title: Cloning and structural analysis of a gene encoding a mouse mastocytoma proteogly

A:Reference number: J00791; MUID:91033033; PMID:2121613

A:Accession: J00791

A:Molecule type: DNA

A:Residues: 1-152 <ANG>

A:Cross-references: GB:M33497

A:Experimental source: mast cell

R:Avraham, S.; Austen, K.F.; Nicodemus, C.F.; Gartner, M.C.; Stevens, R.L.
J. Biol. Chem. 264, 16719-16726, 1989

A:Title: Cloning and characterization of the mouse gene that encodes the peptide core of

A:Reference number: A34399; MUID:89380298; PMID:2777804

A:Accession: A34399

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-152 <AVR>

A:Cross-references: GB:M27293; GB:J05044; NID:9200458; PIDN:AAA39965.1; PID:9387517

R:Kjellen, L.; Petersson, I.; Lillnager, P.; Steen, M.L.; Petersson, U.; Lehtonen, P.;
Blochem. J. 263, 105-113, 1989

A:Title: Primary structure of a mouse mastocytoma proteoglycan core protein.

A:Reference number: S06127; MUID:90104216; PMID:2532501

A:Accession: S06127

A:Molecule type: mRNA

A:Residues: 1-152 <KJE>

A:Cross-references: EMBL:X16133; NID:954085; PIDN:CA34259.1; PID:954086

R:Avraham, S.; Stevens, R.L.; Nicodemus, C.F.; Gartner, M.C.; Austen, K.F.; Weis, J.H.
Proc. Natl. Acad. Sci. U.S.A. 86, 3763-3767, 1989

A:Title: Molecular cloning of a cDNA that encodes the peptide core of a mouse mast cell

A:Reference number: A32779; MUID:89264503; PMID:2726751

A:Accession: A32779

A:Molecule type: mRNA

A:Residues: 1-152 <AV2>

A:Cross-references: GB:J04549; NID:9200973; PIDN:AAA4011.1; PID:9200974

C:Genetics:

A:Gene: SGC

A:Introns: 25/1; 73/2

C:Superfamily: proteoglycan 10K core protein

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-152/Product: proteoglycan 17K core protein #status predicted <MAT>

Query Match 83.7%; Score 77; DB 2; Length 152;
Best Local Similarity 80.0%; Pred. No. 2.8e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPTORARYQWVRCNP 15
DB 26 YPARRARYQWVRCNP 40

RESULT 3

GZRP0

secretory granule proteoglycan core protein precursor - rat

N:Alternate names: 19PG core protein; chondroitin sulfate proteoglycan core protein;

C:Species: Rattus norvegicus (Norway rat)

C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 22-Jun-1999

C:Accession: A25644; A28059; A60653

R:Bourdon, M.A.; Shiga, M.; Ruostehi, E.

J. Biol. Chem. 261, 12534-12537, 1986

A:Title: Identification from cDNA of the precursor form of a chondroitin sulfate prot

A:Reference number: A25644; MUID:86304425; PMID:2427521

A:Accession: A25644

A:Molecule type: mRNA

A:Residues: 1-179 <BOU>

A:Cross-references: GB:K02934; GB:M14282; NID:9207039; PIDN:AAA42171.1; PID:9207040

A:Experimental source: yolk sac carcinoma

R:Avraham, S.; Stevens, R.L.; Gartner, M.C.; Austen, K.F.; Lalley, P.A.; Weis, J.H.
J. Biol. Chem. 263, 7292-7296, 1988

A:Title: Isolation of a cDNA that encodes the peptide core of the secretory granule p

A:Reference number: A28059; MUID:88213412; PMID:3366780

A:Accession: A28059

A:Molecule type: mRNA

A:Residues: 1-179 <AVR>

A:Cross-references: EMBL:U03224; NID:9206108; PIDN:AAA41837.1; PID:9206109

R:Clorda, R.; Chambers, W.H.; Dahl, C.A.; Trucco, M.

Nat. Immun. Cell Growth Regul. 9, 91-102, 1990

A:Title: Isolation and characterization of a cDNA that encodes the core protein of th

A:Reference number: A60653; MUID:90279692; PMID:2352541

A:Accession: A60653

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-179 <GIO>

C:Superfamily: proteoglycan 10K core protein

C:Keywords: chondroitin sulfate proteoglycan; dermatan sulfate; extracellular matrix;

F:1-26/Domain: signal sequence #status predicted <SIG>

F:27-75/Product: activation peptide #status predicted <AP>

F:76-179/Product: proteoglycan 10K core protein #status predicted <MP>

F:90-138/Region: chondroitin sulfate attachment (S-G repeats)

Query Match 83.7%; Score 77; DB 1; Length 179;
Best Local Similarity 80.0%; Pred. No. 3.2e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPTORARYQWVRCNP 15
DB 27 YPARRARYQWVRCNP 41

RESULT 4

T46140

hypothetical protein T3A5.10 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46140

R:Blocker, H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Scheller, C.; Quettler, F.; S

submitted to the Protein Sequence Database, December 1999

A:Reference number: Z23024

A:Accession: T46140

A:Molecule type: preliminary

A>Status: preliminary

A:Residues: 1-209 <BLO>

A:Cross-references: EMBL:AL132979

A:Experimental source: cultivar Columbia; BAC clone T3A5

C:Genetics:

A:Map position: 3

A:Introns: 109/3; 130/2; 186/2

A:Note: T3A5.10

Query Match 48.9%; Score 45; DB 2; Length 209;
Best Local Similarity 50.0%; Pred. No. 6.1;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 PTORARYQWVRCNP 15

Db 196 PLGGGRYEMVKLNP 209

RESULT 5

T06502
hypothetical protein 91 - garden pea chloroplast
C:Species: chloroplast Pisum sativum (garden pea)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: A23041; T06502
R:Rasmussen, O.F.; Stumann, B.M.; Henningsen, K.W.
Nucleic Acids Res. 12, 9143-9153, 1984

A:Title: Nucleotide sequence of a 1.1 kb fragment of the pea chloroplast genome containing
A:Reference number: A23041; MUID:85087925; PMID:6096819
A:Accession: A23041
A:Status: translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 1-91 <RAS>
A:Cross-references: EMBL:X01676; NID:g14204; PIDN:CAA25831.1; PID:g14205
C:Genetics:
A:Genome: chloroplast
C:Keywords: chloroplast

Query Match 47.8%; Score 44; DB 2; Length 91;
Best Local Similarity 41.7%; Pred. No. 4.1;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 PTORARYQWVC 13
|||:||||:
Db 38 PIKKIKIOWISC 49

RESULT 6

H85138
hypothetical protein AT4g12900 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
C:Accession: H85138
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198

A:Accession: H85138
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-231 <STO>
A:Cross-references: GB:NC_001268; NID:g7267992; PIDN:CAB78332.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g12900
A:Map position: 4
C:Superfamily: Arabidopsis thaliana hypothetical protein F7A7.100

Query Match 47.8%; Score 44; DB 2; Length 231;
Best Local Similarity 46.2%; Pred. No. 9.7;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 YPTORARYQWVC 13
:|||||:
Db 110 WPNORLHYKFTIC 122

RESULT 7

T02320
hypothetical protein At2g34320 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F13P17.16
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02320; B84755
R:Rounsley, S.D.; Lin, X.; Kethum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC F13P17 genomic sequence.
A:Reference number: Z14657
A:Accession: T02320

A:Status: translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 1-292 <ROU>

A:Cross-references: EMBL:AC004481; NID:g3337347; PID:g3337363
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84755
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-292 <STO>
A:Cross-references: GB:AE002093; NID:g3337363; PIDN:AC27408.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g34320; F13P17.16
A:Map position: 2

Query Match 47.8%; Score 44; DB 2; Length 292;
Best Local Similarity 65.7%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 8 YQWVRCN 14
|||||:
Db 142 YQWVRCN 148

RESULT 8

D83166
hypothetical protein PA3843 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83166
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L.
.. Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: D83166
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-120 <STO>
A:Cross-references: GB:AE004801; GB:AE004091; NID:g9950011; PIDN:AA07230.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3843

Query Match 46.7%; Score 43; DB 2; Length 120;
Best Local Similarity 53.8%; Pred. No. 7.8;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 TORARYQWVCNP 15
|||:|||||:
Db 103 TQTTSSRWVSCNP 115

RESULT 9

T122274
hypothetical protein F46B3.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T122274
R:Ainscough, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19541
A:Accession: T122274
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-354 <WIL>
A:Cross-references: EMBL:Z81540; PIDN:CAB034398.1; GSPDB:GN00023; CESP:F46B3.9

A:Experimental source: clone F46B3
C:Genetics:

A:Gene: CESP:F46B3.9
A:Map position: 5
A:Introns: 23/1; 55/1; 82/1; 112/1; 142/1; 170/1; 193/1; 223/1; 269/1

Query Match 46.7%; Score 43; DB 2; Length 354;
Best Local Similarity 71.4%; Pred. No. 21;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 9 OMVRCNP 15
11:1111

DB 90 QWTKCNP 96

RESULT 10

H95390
probable reverse transcriptase [imported] - *Sinorhizobium meliloti* (strain 1021) magaple

C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: H95390

R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows

; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, R.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9889, 2001

A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*

A:Reference number: A95262; MUID:21396509; PMID:11481432

A:Accession: H95390

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-505 <KUR>

A:Cross-references: GB:AE006469; PIDN:AAK65690.1; PID:914524181; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA

R:Gallbert, F.; Finan, T.M.; Long, S.R.; Publer, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The complete genome of the legume symbiont *Sinorhizobium meliloti*.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: Sma1875

A:Genome: plasmid

Query Match 46.7%; Score 43; DB 2; Length 505;
Best Local Similarity 58.3%; Pred. No. 29;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 4 ORARYOWRCNP 15
11111111

DB 16 ORKLYOWSKAMP 27

RESULT 11

T29816
hypothetical protein C46A5.9 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T29816
R:Johnson, D.; Stellyes, L.
submitted to the EMBL Data Library, June 1996

A:Description: The sequence of C. elegans cosmid C46A5.
A:Reference number: 220690
A:Accession: T29816
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-717 <JON>
A:Cross-references: EMBL:061948; PIDN:AAB03147.1; GSPDB:GN00022; CESP:C46A5.9
A:Experimental source: strain Bristol N2; clone C46A5
C:Genetics:
A:Gene: CESP:C46A5.9
A:Map position: 4

A:Introns: 11/1; 331/3; 569/3; 671/3

Query Match 46.7%; Score 43; DB 2; Length 717;
Best Local Similarity 46.7%; Pred. No. 40;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 YPQRRARYOWRCNP 15
11111111

DB 57 YELQSTRWEMRRRLNP 71

RESULT 12

T43277
host cell factor 1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T43277

R:Lin, Y.; Hengartner, M.O.; Herr, W.
Mol. Cell. Biol. 19, 909-915, 1999

A:Title: Selected elements of herpes simplex virus accessory factor HCF are highly co

A:Reference number: 22386; MUID:99078028; PMID:9858614

A:Accession: T43277

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-782 <LIID>

A:Cross-references: EMBL:AF072907; NID:q3851533; PIDN:AAD12580.1; PID:q3851534

C:Genetics:

A:Gene: hcf-1

A:Map position: 4

Query Match 46.7%; Score 43; DB 2; Length 782;
Best Local Similarity 46.7%; Pred. No. 43;

Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 1 YPQRRARYOWRCNP 15
11111111

DB 122 YELQSTRWEMRRRLNP 136

RESULT 13

T10198
hypothetical protein F25G13.50 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000

C:Accession: T10198

R:Bevan, M.; Pohl, T.; Weizenecker, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lem

submitted to the Protein Sequence Database, June 1999

A:Reference number: 216991

A:Accession: T10198

A:Molecule type: DNA

A:Residues: 1-233 <BEV>

A:Cross-references: EMBL:AL079349; GSPDB:GN00062; ATSP:F25G13.50

A:Experimental source: cultivar Columbia; BAC clone F25G13

C:Genetics:

A:Gene: ATSP:F25G13.50

A:Map position: 4

Query Match 45.7%; Score 42; DB 2; Length 233;
Best Local Similarity 38.5%; Pred. No. 21;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPQRRARYOWRCNP 13
11111111

DB 105 WPDQKSQYSFIRC 117

RESULT 14

E70951
hypothetical protein RV3199c - *Mycobacterium tuberculosis* (strain H37RV)

C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 26-May-2000


```

C:Accession: E70951
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rejandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70951
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-313 <COL>
A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA1664.1; PID:e124879
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv3199c
C:Superfamily: unassigned mutT domain proteins; mutT domain homology
F:198-233/Domain: mutT domain homology <MUTT>

Query Match          45.7%; Score 42; DB 2; Length 313;
Best Local Similarity 57.1%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 PTQRRARQWRCNP 15
   ||: || | | | |
Db 147 PTKPARAGWSRVNP 160

RESULT 15
B63171
hypothetical protein yidR - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: B63171
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: B63171
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-416 <BLAT>
A:Cross-references: GB:AE000446; GB:U00096; NID:g2367261; PIDN:AACT6712.1; PID:g1790124;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yidR
C:Superfamily: Escherichia coli hypothetical protein yidR

Query Match          45.7%; Score 42; DB 2; Length 416;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 WYRCNP 15
   |||||
Db 302 WYRCNP 307

```

Search completed: April 8, 2003, 11:29:58
Job time : 24.7931 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:28:39 ; Search time 20.6897 Seconds
(without alignments)
44.324 Million cell updates/sec

Title: US-09-496-391-10

Perfect score: 92

Sequence: 1 YPTORARYQWRCNP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	100.0	69	9	US-09-796-692-1379
2	92	100.0	69	9	US-09-796-692-2104
3	92	100.0	69	9	US-09-796-692-2355
4	92	100.0	76	9	US-09-796-692-1346
5	92	100.0	76	9	US-09-796-692-1835
6	92	100.0	76	9	US-09-796-692-2088
7	92	100.0	76	9	US-09-796-692-2318
8	92	100.0	123	9	US-09-796-692-1360
9	92	100.0	123	9	US-09-796-692-2329
10	92	100.0	143	9	US-09-796-692-1372
11	92	100.0	143	9	US-09-796-692-2339
12	92	100.0	148	9	US-09-796-692-1364
13	92	100.0	156	9	US-09-796-692-2357
14	92	100.0	158	9	US-09-796-692-1356
15	92	100.0	158	9	US-09-796-692-2325
16	92	100.0	162	10	US-09-925-300-1362
17	59	64.1	125	9	US-09-796-692-1301
18	59	64.1	125	9	US-09-796-692-2292
19	45	48.9	176	10	US-09-733-507-11

20	41	44.6	43	10	US-09-864-761-39256	Sequence 39256, A
21	40	43.5	66	10	US-09-864-761-48359	Sequence 48359, A
22	40	43.5	813	10	US-09-733-180-7	Sequence 7, Appl
23	40	43.5	844	10	US-09-733-180-2	Sequence 2, Appl
24	39	42.4	49	9	US-09-986-480-296	Sequence 296, App
25	39	42.4	59	9	US-09-986-480-327	Sequence 327, App
26	39	42.4	188	10	US-09-925-301-1331	Sequence 1331, Ap
27	39	42.4	196	10	US-09-733-507-16	Sequence 16, Appl
28	39	42.4	382	9	US-10-078-770-66	Sequence 66, Appl
29	39	42.4	545	9	US-09-978-295A-254	Sequence 254, App
30	39	42.4	545	9	US-09-978-697-254	Sequence 254, App
31	39	42.4	545	9	US-09-978-192A-254	Sequence 254, App
32	39	42.4	545	9	US-09-999-832A-254	Sequence 254, App
33	39	42.4	545	9	US-09-978-189-254	Sequence 254, App
34	39	42.4	545	9	US-10-174-590-58	Sequence 58, Appl
35	39	42.4	545	9	US-10-175-758-58	Sequence 58, Appl
36	39	42.4	545	9	US-10-175-737-58	Sequence 58, Appl
37	39	42.4	545	9	US-10-173-706-58	Sequence 58, Appl
38	39	42.4	545	9	US-10-175-738-58	Sequence 58, Appl
39	39	42.4	545	9	US-10-175-752-58	Sequence 58, Appl
40	39	42.4	545	9	US-10-176-482-58	Sequence 58, Appl
41	39	42.4	545	9	US-10-176-757-58	Sequence 58, Appl
42	39	42.4	545	9	US-10-176-913-58	Sequence 58, Appl
43	39	42.4	545	9	US-10-180-552-58	Sequence 58, Appl
44	39	42.4	545	9	US-10-180-557-58	Sequence 58, Appl
45	39	42.4	545	9	US-10-173-700-58	Sequence 58, Appl

ALIGNMENTS

RESULT 1

US-09-796-692-1379

Sequence 1379, Application US/09796692

Publication No. US20020198362A1

GENERAL INFORMATION:

APPLICANT: Gaiger, Alexander

APPLICANT: Algate, Paul A.

APPLICANT: Manion, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER

FILE REFERENCE: 2077.001200

CURRENT FILING DATE: 2001-03-01

PRIOR FILING DATE: 2000-04-27

PRIOR FILING DATE: 2000-03-01

PRIOR FILING DATE: 2000-03-17

PRIOR FILING DATE: 2000-03-17

PRIOR FILING DATE: 2000-04-27

PRIOR FILING DATE: 2000-05-01

PRIOR FILING DATE: 2000-05-01

PRIOR FILING DATE: 2000-05-04

PRIOR FILING DATE: 2000-05-04

PRIOR FILING DATE: 2000-05-22

PRIOR FILING DATE: 2000-05-22

PRIOR FILING DATE: 2000-07-14

PRIOR FILING DATE: 2000-07-14

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-08-04

PRIOR FILING DATE: 2000-08-04

PRIOR FILING DATE: 2000-08-07

PRIOR FILING DATE: 2000-08-07

NUMBER OF SEQ ID NOS: 9597

SOFTWARE: FASTSEQ for Windows Version 3.0

SEQ ID NO 1379

LENGTH: 69

TYPE: PRT

```
; ORGANISM: Homo sapiens
US-09-796-692-1379

Query Match
Best Local Similarity 100.0%; Score 92; DB 9; Length 69;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPTQARYQWRCNP 15
Db 25 YPTQARYQWRCNP 39

RESULT 2
US-09-796-692-2104
; Sequence 2104, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2104
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2104

Query Match
Best Local Similarity 100.0%; Score 92; DB 9; Length 69;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPTQARYQWRCNP 15
Db 25 YPTQARYQWRCNP 39

RESULT 3
US-09-796-692-2355
; Sequence 2355, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
```

```

; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1346
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1346
```

```

Query Match          100.0%; Score 92; DB 9; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 YPTORARYQWRCNP 15
        |||
Db       32 YPTORARYQWRCNP 46
```

```

RESULT 5
US-09-796-692-1835
; Sequence 1835, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1835
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1835
```

```

Query Match          100.0%; Score 92; DB 9; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 YPTORARYQWRCNP 15
        |||
Db       32 YPTORARYQWRCNP 46
```

```

RESULT 6
US-09-796-692-2088
; Sequence 2088, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2088
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2088

Query Match          100.0%; Score 92; DB 9; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 YPTORARYQWRCNP 15
        |||
Db       32 YPTORARYQWRCNP 46
```

```

RESULT 7
US-09-796-692-2318
; Sequence 2318, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2318
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2318
```

```
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2318
LENGTH: 76
TYPE: PRT
ORGANISM: Homo sapiens
US-09-796-692-2318
```

```
Query Match      100.0%; Score 92; DB 9; Length 76;
Best Local Similarity 100.0%; Pred. No. 1,6e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 YPTORARYQWVRCNP 15
Db 32 YPTORARYQWVRCNP 46
```

```
RESULT 8
US-09-796-692-1360
Sequence 1360, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Galger, Alexander
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
```

```
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1360
LENGTH: 123
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: variant
LOCATION: (1)...(123)
OTHER INFORMATION: Xaa = Any amino acid
US-09-796-692-1360
```

```
Query Match      100.0%; Score 92; DB 9; Length 123;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 YPTORARYQWVRCNP 15
Db 29 YPTORARYQWVRCNP 43
```

```
RESULT 9
US-09-796-692-2329
Sequence 2329, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Galger, Alexander
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
FILE REFERENCE: 2077.001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2329
LENGTH: 123
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
```

```

; NAME/KEY: variant
; LOCATION: (1)...(123)
; OTHER INFORMATION: Xaa - Any amino acid
US-09-796-692-2329

Query Match      100.0%; Score 92; DB 9; Length 123;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 YPTORARYQWRCNP 15
        |||
Db      29 YPTORARYQWRCNP 43

RESULT 10
US-09-796-692-1372
; Sequence 1372, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/228,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1372
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(143)
; OTHER INFORMATION: Xaa - Any amino acid
US-09-796-692-1372

Query Match      100.0%; Score 92; DB 9; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 YPTORARYQWRCNP 15
        |||
Db      29 YPTORARYQWRCNP 43

RESULT 11
US-09-796-692-2339
; Sequence 2339, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077.001200
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2339
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(143)
; OTHER INFORMATION: Xaa - Any amino acid
US-09-796-692-2339

Query Match      100.0%; Score 92; DB 9; Length 143;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 YPTORARYQWRCNP 15
        |||
Db      29 YPTORARYQWRCNP 43

RESULT 12
US-09-796-692-1364
; Sequence 1364, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077.001200
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
```

```

; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1364
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(148)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-796-692-1364
```

```

Query Match      100.0%; Score 92; DB 9; Length 148;
Best Local Similarity 100.0%; Pred. No. 3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 YPTQRAYQWVRGNP 15
Db      32 YPTQRAYQWVRGNP 46
```

```

RESULT 13
US-09-796-692-2357
; Sequence 2357, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
```

```

; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2357
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(156)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-796-692-2357

Query Match      100.0%; Score 92; DB 9; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 YPTQRAYQWVRGNP 15
Db      26 YPTQRAYQWVRGNP 40
```

```

RESULT 14
US-09-796-692-1356
; Sequence 1356, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1356
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(158)
```


; OTHER INFORMATION: Xaa = Any amino acid
US-09-796-692-1356

Query Match

100.0%; Score 92; DB 9; Length 158;

Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTORARYQWVRGNP 15
|||||
DB 28 YPTORARYQWVRGNP 42

RESULT 15

US-09-796-692-2325

; Sequence 2325, Application US/09796692

; Publication No. US20020198362A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander

; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES

; FILE REFERENCE: 2077.001200

; CURRENT APPLICATION NUMBER: US/09/796,692

; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 60/200,303

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,999

; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 60/202,084

; PRIOR FILING DATE: 2000-05-04

; PRIOR APPLICATION NUMBER: 60/206,201

; PRIOR FILING DATE: 2000-05-22

; PRIOR APPLICATION NUMBER: 60/218,950

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: 60/222,903

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: 60/223,416

; PRIOR FILING DATE: 2000-08-04

; PRIOR APPLICATION NUMBER: 60/223,378

; PRIOR FILING DATE: 2000-08-07

; NUMBER OF SEQ ID NOS: 9597

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2325

; LENGTH: 158

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: variant

; LOCATION: (1)...(158)

; OTHER INFORMATION: Xaa = Any amino acid

US-09-796-692-2325

Query Match

100.0%; Score 92; DB 9; Length 158;

Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTORARYQWVRGNP 15
|||||
DB 28 YPTORARYQWVRGNP 42

Search completed: April 8, 2003, 11:52:44
Job time : 20.6897 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:19:38 ; Search time 19.6552 Seconds
(without alignments)
22.454 Million cell updates/sec

Title: US-09-496-391-10
Perfect score: 92
Sequence: 1 YPTORARYQWRCNP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	100.0	158	1	US-07-906-871-10
2	92	100.0	158	1	US-07-906-871-12
3	77	83.7	152	1	US-07-906-871-14
4	70	76.1	16	1	US-07-906-871-4
5	39.5	42.9	2539	4	US-09-413-814-42
6	39	42.4	2205	1	US-08-093-453B-2
7	38	41.3	94	4	US-08-055-006-21
8	37.5	40.8	249	4	US-09-134-001C-3351
9	37.5	40.8	7257	3	US-09-335-409-5
10	37.5	40.8	7257	4	US-09-568-102-5
11	37.5	40.8	7257	4	US-09-567-969-5
12	37.5	40.8	7257	4	US-09-568-480-5
13	37.5	40.8	7257	4	US-09-568-486-5
14	37.5	40.8	7257	4	US-09-567-899-5
15	37.5	40.8	7257	4	US-09-567-899-5
16	37	40.2	127	4	US-09-284-033-6
17	37	40.2	127	4	US-08-729-834B-6
18	37	40.2	289	2	US-08-580-545B-4
19	37	40.2	289	4	US-09-262-653A-4
20	37	40.2	289	4	US-08-867-484A-2
21	37	40.2	397	4	US-09-284-033-4
22	37	40.2	397	4	US-08-729-834B-4
23	37	40.2	398	2	US-08-839-581A-32
24	37	40.2	398	4	US-09-023-591A-32
25	37	40.2	446	1	US-08-665-966-10
26	37	40.2	446	3	US-09-041-780-10
27	37	40.2	467	1	US-08-459-100A-3

28	37	40.2	789	1	US-08-431-080-20	Sequence 20, Appl
29	37	40.2	789	2	US-08-938-534-20	Sequence 20, Appl
30	37	40.2	789	2	US-09-345-294-20	Sequence 20, Appl
31	36.5	39.7	1421	3	US-09-335-409-2	Sequence 2, Appl
32	36.5	39.7	1421	4	US-09-568-102-2	Sequence 2, Appl
33	36.5	39.7	1421	4	US-09-567-969-2	Sequence 2, Appl
34	36.5	39.7	1421	4	US-09-568-480-2	Sequence 2, Appl
35	36.5	39.7	1421	4	US-09-568-486-2	Sequence 2, Appl
36	36.5	39.7	1421	4	US-09-567-899-2	Sequence 2, Appl
37	36.5	39.7	1421	4	US-09-335-409-4	Sequence 2, Appl
38	36.5	39.7	1832	3	US-09-568-102-4	Sequence 4, Appl
39	36.5	39.7	1832	4	US-09-567-969-4	Sequence 4, Appl
40	36.5	39.7	1832	4	US-09-568-480-4	Sequence 4, Appl
41	36.5	39.7	1832	4	US-09-568-486-4	Sequence 4, Appl
42	36.5	39.7	1832	4	US-09-568-486-4	Sequence 4, Appl
43	36.5	39.7	1832	4	US-09-567-899-4	Sequence 4, Appl
44	36.5	39.7	1832	4	US-09-567-899-4	Sequence 4, Appl
45	36.5	39.7	2259	4	US-09-413-814-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-07-906-871-10
Sequence 10, Application US/07906871
Patent No. 5340739
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
APPLICANT: Avraham, Shalom
TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
TITLE OF INVENTION: THREEDEF
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1225 Connecticut Avenue, N.W., Suite 300
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07906,871
FILING DATE: 19920103
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/816,289
FILING DATE: 03 JAN 1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/635,544
FILING DATE: 18 JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/03051
FILING DATE: 13-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/224,035
FILING DATE: 13-JUL-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cimbal, Michele A
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0627.2830004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)833-7533
TELEFAX: (202)833-8716
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: AMINO ACID

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-906-871-10

Query Match 100.0%; Score 92; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTORARYQWVRCNP 15
|||||
DB 28 YPTORARYQWVRCNP 42

RESULT 2
US-07-906-871-12
Sequence 12, Application US/07906871
Patent No. 5340739
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
APPLICANT: Avraham, Shalom
TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1225 Connecticut Avenue, N.W., Suite 300
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/906,871
FILING DATE: 19920103
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/816,289
FILING DATE: 03 JAN 1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/635,544
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/03051
FILING DATE: 13-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/224,035
FILING DATE: 13-JUL-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cimbala, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0627.2830004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)833-7533
TELEFAX: (202)833-8716
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-906-871-12

Query Match 100.0%; Score 92; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 YPTORARYQWVRCNP 15
|||||

DB 28 YPTORARYQWVRCNP 42

RESULT 3

US-07-906-871-14
Sequence 14, Application US/07906871
Patent No. 5340739
GENERAL INFORMATION:

APPLICANT: Stevens, Richard L.
APPLICANT: Avraham, Shalom
TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1225 Connecticut Avenue, N.W., Suite 300
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/906,871
FILING DATE: 19920103
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/816,289
FILING DATE: 03 JAN 1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/635,544
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/03051
FILING DATE: 13-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/224,035
FILING DATE: 13-JUL-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cimbala, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0627.2830004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)833-7533
TELEFAX: (202)833-8716
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-906-871-14

Query Match 83.7%; Score 77; DB 1; Length 152;
Best Local Similarity 80.0%; Pred. No. 4.2e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPTORARYQWVRCNP 15
||:|||||
DB 26 YPARARYQWVRCNP 40

RESULT 4
US-07-906-871-4
Sequence 4, Application US/07906871
Patent No. 5340739
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
APPLICANT: Avraham, Shalom

;; TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
;; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
;; FILE OF INVENTION: THEREOF
;; NUMBER OF SEQUENCES: 18
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;; STREET: 1225 Connecticut Avenue, N.W., Suite 300
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/906,871
;; FILING DATE: 19920103
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/816,289
;; FILING DATE: 03 JAN 1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/635,544
;; FILING DATE: 18-JAN-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US89/03051
;; FILING DATE: 13-JUL-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/224,035
;; FILING DATE: 13-JUL-1988
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cimbala, Michele A
;; REGISTRATION NUMBER: 33,851
;; REFERENCE/DOCKET NUMBER: 0627, 2830004
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)833-8716
;; TELEFAX: (202)833-7533
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 16 amino acids
;; TYPE: AMINO ACID
;; STRANDEDNESS: both
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-07-906-871-4

Query Match 76.1%; Score 70; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 5.5e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPTORARYQWVR 12
DB 5 YPTORARYQWVR 16

RESULT 5
US-09-413-814-42
Sequence 42, Application US/09413814
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloecker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoffe, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans

;; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
;; TITLE OF INVENTION: heteropolypeptide compounds
;; FILE REFERENCE: PCT/US 99/23535
;; CURRENT APPLICATION NUMBER: US/09/413,814
;; CURRENT FILING DATE: 1999-10-07
;; EARLIER FILING DATE: 1998-10-09
;; NUMBER OF SEQ ID NOS: 107
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 42
;; LENGTH: 2539
;; TYPE: PRT
;; ORGANISM: Sorangium cellulosum
;; US-09-413-814-42

Query Match 42.9%; Score 39.5; DB 4; Length 2539;
Best Local Similarity 46.7%; Pred. No. 5.6e+02;
Matches 7; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

OY 1 YPTORARYQWVRCNP 15
DB 1356 YPFERRRF-WIEGNP 1369

RESULT 6
US-08-093-453B-2
Sequence 2, Application US/08093453B
Patent No. 5439814
GENERAL INFORMATION:
APPLICANT: Frey, Terry K.
APPLICANT: Dominguez, Geraldina
APPLICANT: Wang, Chin Yen
TITLE OF INVENTION: Modified Infectious Rubella Virus
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jamie L. Greene, Jones & Askew
STREET: 191 Peachtree Street, 37th Floor
CITY: Atlanta
STATE: Georgia
COUNTRY: United States
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50
COMPUTER: Macintosh
OPERATING SYSTEM: 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/093,453B
FILING DATE: 19 JUL 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 07/722,334
FILING DATE: 28 JUN 1991
ATTORNEY/AGENT INFORMATION:
NAME: Greene, Jamie L.
REGISTRATION NUMBER: 32,467
REFERENCE/DOCKET NUMBER: 07362-0101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-818-3700
TELEFAX: 404-818-3799
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2205 amino acids
TYPE: amino acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE: Rubella virus
ORGANISM: Rubella virus
STRAIN: Thierlen

US-08-093-453B-2

Query Match 42.4%; Score 39; DB 1; Length 2205;
Best Local Similarity 50.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 YPTQARYQW 10

DB 1462 WPTERSHWT 1471

RESULT 7

US-08-055-006-21

; Sequence 21, Application US/08055006
; Patent No. 6464978

; GENERAL INFORMATION:

; APPLICANT: Brostloff, Steven W.

; Wilson, Darcy B.

; Smith, Lawrence R.

; Gold, Daniel P.

; Carlo, Dennis J.

; TITLE OF INVENTION: Vaccination and Methods Against Multiple

; Sclerosis Resulting From Pathogenic Responses By Specific T

; Cell Populations

; NUMBER OF SEQUENCES: 58

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CAMPBELL & FLORES LLP

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: United States

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION NUMBER: US/08/055,006

; APPLICATION NUMBER: US/08/055,006

; FILING DATE: 09-Feb-1993

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-1M 9611

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619)535-9001

; TELEFAX: (619)535-8949

; INFORMATION FOR SEQ ID NO: 21:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 94 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 21:

US-08-055-006-21

QY 3 TORARYQWVRCNP 15

DB 12 TEKKRYQDLTCDP 24

RESULT 8

US-09-134-001C-3351

; Sequence 3351, Application US/09134001C
; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3351

; LENGTH: 249

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3351

QY 1 YPTQARYQWVRCNP 15

DB 113 YQTFQPAQYTLVTCNP 128

RESULT 9

US-09-335-409-5

; Sequence 5, Application US/09335409
; Patent No. 6121029

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/335,409

; CURRENT FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 5

; LENGTH: 7257

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-335-409-5

QY 1 YPTQARYQWV 11

DB 5970 YPMQRERY-WV 5979

RESULT 10

US-09-568-102-5

; Sequence 5, Application US/09568102
; Patent No. 6346404

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,102

; CURRENT FILING DATE: 2000-05-10

; PRIOR APPLICATION NUMBER: 09/335,409

; PRIOR FILING DATE: 1999-06-17

```
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-5
```

```
Query Match          40.8%; Score 37.5; DB 4; Length 7257;
Best Local Similarity 72.7%; Pred. No. 3.3e+03;
Matches 8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

```
OY      1 YPTORARYQW 11
      11111111
Db 5970 YPMQREXY-WV 5979
```

```
RESULT 11
US-09-567-969-5
; Sequence 5, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligou, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-969-5
```

```
Query Match          40.8%; Score 37.5; DB 4; Length 7257;
Best Local Similarity 72.7%; Pred. No. 3.3e+03;
Matches 8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

```
OY      1 YPTORARYQW 11
      11111111
Db 5970 YPMQREXY-WV 5979
```

```
RESULT 12
US-09-568-480-5
; Sequence 5, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligou, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
```

```
; ORGANISM: Sorangium cellulosum
US-09-568-480-5
```

```
Query Match          40.8%; Score 37.5; DB 4; Length 7257;
Best Local Similarity 72.7%; Pred. No. 3.3e+03;
Matches 8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

```
OY      1 YPTORARYQW 11
      11111111
Db 5970 YPMQREXY-WV 5979
```

```
RESULT 13
US-09-568-486-5
; Sequence 5, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligou, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-5
```

```
Query Match          40.8%; Score 37.5; DB 4; Length 7257;
Best Local Similarity 72.7%; Pred. No. 3.3e+03;
Matches 8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

```
OY      1 YPTORARYQW 11
      11111111
Db 5970 YPMQREXY-WV 5979
```

```
RESULT 14
US-09-568-472-5
; Sequence 5, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligou, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-5
```

```
Query Match          40.8%; Score 37.5; DB 4; Length 7257;
Best Local Similarity 72.7%; Pred. No. 3.3e+03;
```

Matches 8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 YPTORARYQWV 11
11111111
Db 5970 YPWQRRRY-WV 5979

RESULT 15
US-09-567-899-5
; Sequence 5, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-5

Query Match 40.8%; Score 37.5; DB 4; Length 7257;
Best Local Similarity 72.7%; Pred. No. 3.3e+03;
Matches 8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 YPTORARYQWV 11
11111111
Db 5970 YPWQRRRY-WV 5979

Search completed: April 8, 2003, 11:31:18
Job time : 21.6552 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:17:38 ; Search time 53.2759 seconds
(without alignments)
58.013 Million cell updates/sec

Title: US-09-496-391-9
Perfect score: 90
Sequence: 1 YPARRARYQWVRCRP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	46	51.1	351	16	092RH6
2	45	50.0	336	2	054077
3	45	50.0	415	16	092RH6
4	44	48.9	111	10	Q39682
5	44	48.9	183	16	08RXX7
6	44	48.9	328	3	09Y701
7	43	47.8	296	10	08S200
8	43	47.8	1039	10	08W3F7
9	43	47.8	1708	17	026769
10	42.5	46.7	7576	2	092G44
11	42	46.7	83	16	09R127
12	42	46.7	119	17	08XY52
13	42	46.7	177	17	08RXX7
14	42	46.7	195	10	09FX90
15	42	46.7	195	10	094CL9
16	42	46.7	309	15	091G01

17	42	46.7	332	5	096916	096916 anophles g
18	42	46.7	382	2	09LCU2	091C12 pseudomonas
19	42	46.7	390	2	068707	091C12 yersinia pe
20	42	46.7	735	12	09DUC9	09DUC9 tt virus. o
21	42	46.7	785	5	09V089	09V089 drosophila
22	42	46.7	867	5	095RC2	095RC2 drosophila
23	42	46.7	956	5	09GRN3	09GRN3 leishmania
24	42	46.7	2861	5	09ULC3	09ULC3 leishmania
25	41.5	46.1	7525	2	09IEB0	09IEB0 streptomyces
26	41	45.6	192	10	09FS28	09FS28 pismu sativ
27	41	45.6	204	16	091AB9	091AB9 pseudomonas
28	41	45.6	222	10	09FKB5	09FKB5 arabidopsis
29	41	45.6	346	4	08W015	08W015 homo sapien
30	41	45.6	360	4	09B0M1	09B0M1 homo sapien
31	41	45.6	398	5	P91264	P91264 caenorhabdi
32	41	45.6	458	17	09VCR2	09VCR2 aerypyrum p
33	41	45.6	645	4	09HB94	09HB94 homo sapien
34	41	45.6	2218	12	009705	009705 lassa virus
35	41	45.6	2703	10	09MAZ4	09MAZ4 arabidopsis
36	41	45.6	2810	10	09FKS4	09FKS4 arabidopsis
37	40.5	45.0	484	17	08TWJ9	08TWJ9 methanopyru
38	40.5	45.0	568	5	09VSK8	09VSK8 drosophila
39	40	44.4	129	5	097048	097048 pinctada fu
40	40	44.4	129	5	090AH3	090AH3 pinctada fu
41	40	44.4	131	5	09TW98	09TW98 pinctada fu
42	40	44.4	131	5	09TW92	09TW92 pinctada fu
43	40	44.4	132	12	08ORX6	08ORX6 chimpanzee
44	40	44.4	287	10	08RU18	08RU18 oryza sativ
45	40	44.4	372	2	007846	007846 raietonia s

ALIGNMENTS

RESULT 1

ID	Q92RH6	PREDIMINARY:	PRT:	351 AA.
AC	Q92RH6	01-DEC-2001 (TREMUREL 19, Created)		
DT	01-DEC-2001 (TREMUREL 19, Last sequence update)			
DT	01-MAR-2002 (TREMUREL 20, Last annotation update)			
DE	Putative chaperone protein.			
GN	R00897 OR SMC00003.			
OS	Rhizobium meliloti (Sinorhizobium meliloti).			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.			
CC	Rhizobiaceae; Sinorhizobium.			
OX	NCBI_TaxID=382;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=1021;			
RX	MEDLINE=21396507; PubMed=11481430;			
RA	Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,			
RA	Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,			
RA	Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,			
RA	Pohl T., Portetelle D., Puehler A., Purrelle B., Ramepferger U.,			
RA	Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,			
RT	"Analysis of the chromosome sequence of the legume symbiont			
RT	Sinorhizobium meliloti strain 1021."			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).			
DR	EMBL; AL591785; CAC45469.1;			
DR	InterPro; IPR002939; DnaJ_C.			
DR	InterPro; IPR001623; DnaJ_N.			
DR	Pfam; PF00226; DnaJ_1.			
DR	Pfam; PF01556; DnaJ_C_1.			
DR	PROSITE; PS00636; DnaJ_1; UNKNOWN_1.			
DR	PROSITE; PS00076; DnaJ_2; 1.			
KW	Complete proteome.			
SO	SEQUENCE 351 AA; 38264 MW; 286C201EC60EAA6A0 CRC64;			
Query Match	51.1%; Score 46; DB 16; Length 351;			
Best Local Similarity	63.6%; Pred. No. 14;			
Matches	7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;			

```

OY 2 PARRAYQWVRC 12
DB 62 PVRASRYDMAR 72

RESULT 2
OY 054077 PRELIMINARY; PRT; 336 AA.
AC 054077:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE Hypothetical 37.1 kDa protein.
ORF.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Plasmid pSUP202.79.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OX Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 8253;
RX MEDLINE=98296071; PubMed=9630474;
RA Kort R., Phillips-Jones M.K., Van Aalten D.M.F., Haker A.,
RA Hoffer S.M., Hellinger K.J., Crieleard W.;
RT "Sequence, chromophore extraction and 3-D model of the photoactive
RT yellow protein from Rhodobacter sphaeroides."
RL Biochim. Biophys. Acta 1385:1-6(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 8253;
RX MEDLINE=20389611; PubMed=10930745;
RA Duggan P.S., Parker S.D., Phillips-Jones M.K.;
RT "Characterisation of a Rhodobacter sphaeroides gene that encodes a
RT product resembling Echerichia coli cytochrome b561 and R. sphaeroides
RT cytochrome b562."
RL FEMS Microbiol. Lett. 189:239-246(2000).
DR EMBL; AJ002398; CA05382.1;
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 336 AA; 37075 MW; 67853A9960315D19 CRC64;

Query Match
Best Local Similarity 50.0%; Score 45; DB 2; Length 336;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 PARRAYQWVRC 15
DB 219 PCRAPPMPRCPP 232

RESULT 3
OY 092P21 PRELIMINARY; PRT; 415 AA.
AC 092P21:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Putative transport transmembrane protein.
GN R01579 OR SMC01212.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Anpe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreaño S., Gloux S.,
RA Godt T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
RA Pohl T., Portetle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galbert F.;
RT "Analysis of the chromosome sequence of the legume symbiont

```

```

RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591787; CAC46158.1;
DR InterPro; IPR000566; Lipocin_cyFABP.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tf_1.
DR PROSITE; PS00213; LIPOLALIN; UNKNOWN_1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 415 AA; 43992 MW; 8C22BFBEBE9314AC2 CRC64;

Query Match
Best Local Similarity 50.0%; Score 45; DB 16; Length 415;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 PARRAYQWVRC 15
DB 193 PANRRFEWHRANP 206

RESULT 4
OY 039682 PRELIMINARY; PRT; 111 AA.
AC 039682:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE Glycine-rich protein (Fragment).
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DANVER HALF-LONG; TISSUE-SOMATIC EMBRYOS AT THE GLOBAL STAGE;
RA Lin X., Huang G.-J., Zimmerman J.L.;
RT "Isolation and characterization of a diverse set of genes from carrot
RT somatic embryos."
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U47097; AAB01097.1;
FT NON-TER
SQ SEQUENCE 111 AA; 13829 MW; 082789CADAFA1C2A CRC64;

Query Match
Best Local Similarity 48.9%; Score 44; DB 10; Length 111;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 PARRAYQWVRC 14
DB 35 PPRGRMCWRCR 47

RESULT 5
OY 08RBX7 PRELIMINARY; PRT; 183 AA.
AC 08RBX7:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Hypothetical protein TTE0679.
GN TTE0679.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB47 / JCM11007;
RX MEDLINE=2192816; PubMed=1197336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen K., Wang J., Yu J., Yang H.;

```



```

DE Magnesium chelatase subunit.
GN MTH673.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Sailer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE008447; AAB85178.1; -.
DR InterPro: IPR003672; COBN/Mg_chlase.
DR InterPro: IPR000183; Decarboxylase2.
DR Pfam: PF02514; COBN-Mg_chel. 1.
DR PROSITE: PS00678; ODR_DC_2.1; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 1708 AA; 187987 MW; 59E0A6B115ACC00E CRC64;

Query Match 47.8%; Score 43; DB 17; Length 1708;
Best Local Similarity 46.2%; Pred. No. 26+02;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 YPARRAYQWVRC 13
DB 192 YPGNTALHOMIDC 204

RESULT 10
O9ZG44 PRELIMINARY; PRT; 7576 AA.
AC O9ZG44;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE FK506 polyketide synthase.
GN FKBB.
OS Streptomyces sp. MA6548.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=82632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MA6548;
RX MEDLINE=98451508; PubMed=9780228;
RA Motamedi H., Shafiee A.;
RT "The biosynthetic gene cluster for the macrolactone ring of the
RT immunosuppressant FK506.";
RL Eur. J. Biochem. 256:528-534(1998).
DR EMBL; AF082100; AAC68815.1; -.
DR HSSP: P08659; ILCT.
DR InterPro: IPR001227; AC_transferase.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR000292; For/Nlt_transpt.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR00794; Ketoacyl-synt.
DR InterPro: IPR003880; Ppantate_attach.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00698; Acyl_transf. 4.
DR Pfam: PF00107; adh_zinc. 1.
DR Pfam: PF00501; AMP-binding. 1.
DR Pfam: PF00109; ketoacyl-synt. 4.
DR Pfam: PF02801; ketoacyl-synt_C. 4.

```

```

DR Pfam: PF00550; pp-binding. 5.
DR PROSITE: PS50075; ACP_DOMAIN; 5.
DR PROSITE: PS00455; AMP_BINDING; 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 4.
DR PROSITE: PS01005; FORMATE_NITRITE_TP_1; UNKNOWN_1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Phosphopantetheine transferase.
SQ SEQUENCE 7576 AA; 790105 MW; 49765AB9EE66DD96 CRC64;

Query Match 47.2%; Score 42.5; DB 2; Length 7576;
Best Local Similarity 60.0%; Pred. No. 1e+03;
Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 YPARRAYQWVRC 15
DB 3620 YPFRNRDXY-WVDAEP 3633

RESULT 11
O9RI27 PRELIMINARY; PRT; 83 AA.
AC O9RI27;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE YPCD1.03c protein (Fragment).
GN YPCD1.03c.
OS Yersinia pestis.
OC Plasmid pCD1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS; PLASMID=PCD1;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tlball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skellon J., Stevens K., Whitehead S., Barrrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AL117189; CAB54880.1; -.
KW Plasmid; Complete proteome.
FT NON_TER 83
SQ SEQUENCE 83 AA; 9943 MW; A0B363D58202C3F8 CRC64;

Query Match 46.7%; Score 42; DB 16; Length 83;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 PARRARYQWVRC 15
DB 44 PARTRHKWKLKP 57

RESULT 12
O8XY52 PRELIMINARY; PRT; 119 AA.
AC O8XY52;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative DNA-methyltransferase protein (EC 2.1.1.-).
GN RSC1911 OR RS03480.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.

```

OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,
 Ailat M., Billault A., Brothier P., Camus J.C., Catalico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunac N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sguier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*.";
 RL Nature 415:497-502(2002).
 DR EMBL; AL646067; CAD15613.1; -;
 KW Transferase; Methyltransferase; Complete proteome.
 SQ SEQUENCE 119 AA; 13347 MW; FFA4B1678DAF3F8B CRC64;

Query Match 46.7%; Score 42; DB 16; Length 119;
 Best Local Similarity 50.0%; Pred. No. 21;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 ARRARYOWRCK 14
 Db 62 AKERHOMGCE 73

RESULT 13

ID 08TNA7 PRELIMINARY; PRT; 177 AA.
 AC 08TNA7;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Predicted protein.
 GN MA2386.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Althor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McEwen P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kueltner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.O.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of *Methanosarcina acetivorans* reveals extensive metabolic
 RT and physiological diversity";
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AEO10929; AAM05772.1; -;
 KW Complete proteome.
 SQ SEQUENCE 177 AA; 20878 MW; CAD2FF290DD9F51 CRC64;

Query Match 46.7%; Score 42; DB 17; Length 177;
 Best Local Similarity 38.5%; Pred. No. 31;
 Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 YPARRARYOWRCK 13
 Db 115 YDSEKNOYKWLRC 127

RESULT 14
 Q9FX90 PRELIMINARY; PRT; 195 AA.

AC Q9FX90;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE F14J22.14 protein.
 GN F14J22.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids; II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Federici N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Alrafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
 RA Gonzalez A., Kremetska I., Kim C., Lenz C., Li J., Liu S.,
 RA Luros S., Schwartz J., Shinn P., Tortum M., Vysotskaia V.S.,
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC011807; AAC13048.1; -;
 SQ SEQUENCE 195 AA; 21965 MW; 6A7E0867B752E5FF CRC64;

Query Match 46.7%; Score 42; DB 10; Length 195;
 Best Local Similarity 77.8%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 7 RYOWWSLKP 15
 Db 187 RYOWWSLKP 195

RESULT 15

ID 09AC19 PRELIMINARY; PRT; 195 AA.
 AC 09AC19;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Cyclin-dependent kinase inhibitor 7.
 GN KRP7.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids; II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21342510; PubMed=11449057;
 RA de Veylder L., Beeckman T., Beeckman G.T.S., Krols L., Terras F.,
 RA Landrieu I., Van der Schueren E., Maes S., Naudts M., Inze D.;
 RT "Functional analysis of Cyclin-dependent kinase inhibitors of
 RT Arabidopsis";
 RL Plant Cell 13:1653-1668(2001).
 DR EMBL; AJ301558; CAC41621.1; -;
 KW Kinase.
 SQ SEQUENCE 195 AA; 22008 MW; 60B81B47B52F6DF CRC64;

Query Match 46.7%; Score 42; DB 10; Length 195;
 Best Local Similarity 77.8%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 7 RYOWWSLKP 15
 Db 187 RYOWWSLKP 195

Search completed: April 8, 2003, 11:28:18
 Job time : 56.2759 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:12:02 ; Search time 11.6379 Seconds

(Without alignments)
53.458 Million cell updates/sec

Title: US-09-496-391-9
Perfect score: 90
Sequence: 1 YPARRRARYQWVRCRP 15

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	90	100.0	152	1	PGSG_MOUSE
2	90	100.0	179	1	PGSG_RAT
3	77	85.6	158	1	PGSG_HUMAN
4	42	46.7	315	1	ISTA_SHISO
5	42	46.7	390	1	ISTA_PSEAE
6	40	44.4	394	1	EK12_HUMAN
7	39.5	43.9	3898	1	POLG_HCYA
8	39.5	43.9	3898	1	POLG_HCYA
9	39	43.3	398	1	Y4QJ_RHISN
10	39	43.3	2471	1	NTC2_RAT
11	39	43.3	3649	1	ACVS_NOCIA
12	38.5	42.8	203	1	RL5B_YEAST
13	38.5	42.8	508	1	MM19_HUMAN
14	38	42.2	150	1	YMBW_YEAST
15	38	42.2	307	1	RNHL_YEAST
16	38	42.2	517	1	EAS_DROME
17	38	42.2	2373	1	CCAH_HUMAN
18	37.5	41.7	203	1	RL5A_YEAST
19	37	41.1	67	1	YPE1_NPVLD
20	37	41.1	286	1	PPG_STRCO
21	37	41.1	333	1	PDXA_CAUCR
22	37	41.1	452	1	EK11_HUMAN
23	37	41.1	452	1	PROX_MYCTU
24	37	41.1	568	1	YBAE_ECOLI
25	37	41.1	859	1	NIA_PICAN
26	37	41.1	861	1	ORC1_HUMAN
27	37	41.1	1516	1	Y819_PSEAE
28	36.5	40.6	54	1	RL37_THEAC
29	36.5	40.6	54	1	RL37_THEYO
30	36.5	40.6	92	1	RL37_BOVIN
31	36.5	40.6	96	1	RL37_HUMAN
32	36.5	40.6	96	1	RL37_ICYPU
33	36.5	40.6	581	1	PLRL_CEREL

34	36.5	40.6	616	1	PLRL_RABIT	P14787 oryctolagus
35	36.5	40.6	2715	1	TRX2_HUMAN	Q9umh6 homo sapien
36	36	40.0	92	1	HRX_MOUSE	P70678 mus musculu
37	36	40.0	112	1	VE4_BPV2	P11301 bovine papi
38	36	40.0	121	1	Y162_BURCE	P24576 burkholderi
39	36	40.0	143	1	Y880_MYCLE	Q33060 mycobacteri
40	36	40.0	143	1	Y880_MYCLE	Q10542 mycobacteri
41	36	40.0	200	1	REP2_ZYGFE	P13742 zygosacchar
42	36	40.0	335	1	NU2M_LUMFE	Q34951 lumbricus t
43	36	40.0	365	1	VSQP_EBOIC	Q66811 ebola virus
44	36	40.0	416	1	Y1DR_ECOLI	P31455 escherichia
45	36	40.0	521	1	VGLC_HSVBC	P14378 bovine herp

ALIGNMENTS

RESULT 1	ID	PGSG_MOUSE	STANDARD	PRT	152 AA.
AC	PGSG_MOUSE				
AC	P13609:				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	Secretory granule proteoglycan core protein precursor (Mastocytoma				
DE	proteoglycan core protein) (Serglycin).				
GN	PRG1 OR PRG.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_Taxid=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=89264503; PubMed=2726751;				
RA	Avraham S., Stevens R.L., Nicodemus C.F., Gartner M.C., Austen K.F.,				
RA	Wels J.H.;				
RT	"Molecular cloning of a cDNA that encodes the peptide core of a mouse				
RT	mast cell secretory granule proteoglycan and comparison with the				
RT	analogous rat and human cDNA.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 86:3763-3767(1989).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=89380298; PubMed=2777804;				
RA	Avraham S., Austen K.F., Nicodemus C.F., Gartner M.C., Stevens R.L.;				
RA	Petersson U., Lehtonen P., Karlsson T., Ruoslahti E., Hellman L.;				
RT	"Primary structure of a mouse mastocytoma proteoglycan core protein.";				
RT	peptide core of secretory granule proteoglycans and expression of				
RT	this gene in transfected rat-1 fibroblasts.";				
RL	J. Biol. Chem. 264:16719-16726(1989).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90104216; PubMed=2532501;				
RA	Kjellen L., Petersson U., Lillhager P., Steen M.L.,				
RA	Petersson U., Lehtonen P., Karlsson T., Ruoslahti E., Hellman L.;				
RT	"Primary structure of a mouse mastocytoma proteoglycan core protein.";				
RT	Biochem. J. 263:105-113(1989).				
RL	[4]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=9103033; PubMed=2121613;				
RA	Angerth T., Huang R., Aveskogh M., Petersson U., Kjellen L.,				
RA	Hellman L.;				
RT	"Cloning and structural analysis of a gene encoding a mouse				
RT	mastocytoma proteoglycan core protein; analysis of its evolutionary				
RT	relation to three cross hybridizing regions in the mouse genome.";				
RL	Gene 93:235-240(1990).				
RN	[5]				
RP	SEQUENCE FROM N.A.				
RA	Elliot J.F., Pohajdak B., Talbot D., Miller C.L., Helgason C.D.,				
RA	Blaesley R.C., Paetkau V.H.;				
RL	Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.				
CC	-i- FUNCTION: CORE PROTEIN FOR HIGHLY ACIDIC PROTEOGLYCAN CONTAINING				
CC	GLYCOSAMINOGLYCAN THAT ARE ALMOST EXCLUSIVELY CHONDROITIN SULFATE				
CC	E.				
CC	-i- PTM: THE PEPTIDE CORE OF THE MATURE MOLECULE STORED IN THE				

```

CC CC SECRETORY GRANULE IS ONLY 10 KDa. THIS INDICATES THAT FURTHER
CC CC DEGRADATION OF THE 129 AA PRO FORM OF THE PEPTIDE CORE OCCURS AT
CC CC THE N- AND/OR AT ITS C-TERMINUS.
CC CC
CC CC -1- SIMILARITY: TO CORRESPONDING PROTEINS IN HUMAN AND RAT.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: M27393; AAA39965.1; -.
CC CC DR EMBL: M27391; AAA39965.1; JOINED.
CC CC DR EMBL: M27392; AAA39965.1; JOINED.
CC CC DR EMBL: J04549; AAA40111.1; -.
CC CC DR EMBL: X16133; CAA34259.1; -.
CC CC DR EMBL: M34603; AAA39991.1; -.
CC CC DR EMBL: M33499; AAA39900.1; -.
CC CC PIR: J00791; J00791.
CC CC MGD: MGI:97756; PTG.
CC CC KW Glycoprotein; Proteoglycan; Repeat; Signal.
CC CC FT SIGNAL 1 25
CC CC FT CHAIN 26 152
CC CC FT DOMAIN 89 108 SECRETORY GRANULE PROTEOGLYCAN CORE
CC CC FT DISULFID 38 47 PROTEIN.
CC CC FT SEQUENCE 152 AA; 16711 MW; DEC9829BA31036F CRC64;
CC CC
CC CC Query Match 100.0%; Score 90; DB 1; Length 152;
CC CC Best Local Similarity 100.0%; Pred. No. 5.1e-08;
CC CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
QY 1 YPARRAYQWVWCKP 15
QY |||||||||||||||
DB 26 YPARRAYQWVWCKP 40

```

```

RA MEDLINE=90276692; PubMed=2352541.
RX Giorda R., Chambers W.H., Dahl C.A., Trucco M.;
RT "Isolation and characterization of a cDNA that encodes the core
RT protein of the cytolytic granule proteoglycan in rat natural killer
RT cells."
RT Nat. Immun. Cell Growth Regul. 9:91-102(1990).
RN [4]
RP SEQUENCE OF 75-179 FROM N.A.
RX MEDLINE=85140287; PubMed=3919394;
RA Bourdon M.A., Oldberg A., Pleschbacher M.D., Ruoslahti E.;
RT "Molecular cloning and sequence analysis of a chondroitin sulfate
RT proteoglycan cDNA."
CC Proc. Natl. Acad. Sci. U.S.A. 82:1321-1325(1985).
CC -I- SIMILARITY: TO CORRESPONDING PROTEINS IN HUMAN AND MOUSE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02934; AAA42171.1; -.
DR EMBL; J03224; AAA41837.1; -.
DR PIR; A25644; GZPTO.
DR PIR; A28059; A28059.
DR PIR; A60653; A60653.
KW Glycoprotein; Proteoglycan; Repeat; Signal.
FT SIGNAL 1 26
FT PROPEP 27 75 ACTIVATION PEPTIDE.
FT CHAIN 76 179 SECRETORY GRANULE PROTEOGLYCAN CORE
FT FT PROTEIN.
FT DOMAIN 90 137 24 X 2 AA TANDEM REPEATS OF S-G.
FT DISULFID 39 48 POTENTIAL.
SQ SEQUENCE 179 AA; 18577 MW; D2E2A8E7D3AAAD06 CRC64;

Query Match 100.0%; Score 90; DB 1; Length 179;
Best Local Similarity 100.0%; Pred. No. 6e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPARRARYOWRCRP 15
DB 27 YPARRARYOWRCRP 41

RESULT 3
PGSC_HUMAN STANDARD; PRT; 158 AA.
AC P10124;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Secretory granule proteoglycan core protein precursor (platelet
DE proteoglycan core protein) (P.Pg) (hematopoietic proteoglycan core
DE protein) (Serglycin).
DE PRG1 OR PRG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90016819; PubMed=2798108;
RA Stellegreht C.M., Saunders G.F.;
RT "Nucleotide sequence of a cDNA encoding a hemopoietic proteoglycan
RT core protein."
RL Nucleic Acids Res. 17:7523-7523(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88213411; PubMed=2835370;
RA Stevens R.L., Avraham S., Gartner M.C., Bruns G.A.P., Austen K.F.,
RA Weiss J.H.;
```


RT "Isolation and characterization of a cDNA that encodes the peptide
RT core of the secretory granule proteoglycan of human promyelocytic
RT leukemia HL-60 cells."
RL J. Biol. Chem. 263:7287-7291(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90202841; Pubmed=2180935;
RA Nicodemus C.F., Avraham S., Austen K.F., Purdy S., Jablonski J.,
RA Stevens R.L.;
RT "Characterization of the human gene that encodes the peptide core of
RT secretory granule proteoglycans in promyelocytic leukemia HL-60 cells
RT and analysis of the translated product."
RL J. Biol. Chem. 263:5889-5896(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 28-158, AND SEQUENCE OF 34-158 FROM N.A.
RX MEDLINE=88296856; Pubmed=3402609;
RA Alliel P.M., Perin J.-P., Maillet P., Bonnet F., Rosa J.-P.,
RA Jolles P.;
RT "Complete amino acid sequence of a human platelet proteoglycan."
RL FEBS Lett. 236:123-126(1988).
RN [6]
RP SEQUENCE OF 28-93.
RX MEDLINE=89104992; Pubmed=3214420;
RA Perin J.-P., Bonnet F., Maillet P., Jolles P.;
RT "Characterization and N-terminal sequence of human platelet
RT proteoglycan."
RL Biochem. J. 255:1007-1013(1988).
CC -1- FUNCTION: MAY NEUTRALIZE HYDROLYTIC ENZYMES.
CC -1- INDUCTION: BY BBV.
CC -1- SIMILARITY: TO CORRESPONDING PROTEINS IN RAT AND MOUSE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M33651; AAA60322.1; -
DR EMBL: M33649; AAA60322.1; JOINED.
DR EMBL: M33650; AAA60322.1; JOINED.
DR EMBL: X17042; CAA34900.1; -
DR EMBL: X12765; CAA31255.1; -
DR EMBL: BC015516; AAH15516.1; -
DR EMBL: J03223; AAA60179.1; -
DR PIR: A28058; A28058.
DR PIR: A35183; A35183.
DR PIR: S01126; S01126.
DR PIR: S01819; S01819.
DR PIR: S09610; S09610.
DR Genew: HGNC:9361; PRG1.
DR MTM: 177040; -
KM Glycoprotein; Proteoglycan; Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 158
FT
FT DOMAIN 94 111
FT CARBOHYD 94 94
FT FT 96 96
FT DISULFID 40 49
FT CONFLICT 139 139
SQ SEQUENCE 158 AA; 17624 MW; 1275B7F39FE91476 CRC64;
Query Match 85.6%; Score 77; DB 1; Length 158;
Best Local Similarity 80.0%; Pred. No. 7.3e-06;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YPARRARYWVRCNP 15
Db 28 YPTQFARARYWVRCNP 42
RESULT 4
ID ISTA_SHISO STANDARD; PRT; 315 AA.
AC P16944;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase for insertion sequence element IS640.
GN IS7A.
OS Shigella sonnei.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=624;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88062685; Pubmed=2824781;
RA Matsutani S., Ohtsubo H., Maeda Y., Ohtsubo E.;
RT "Isolation and characterization of IS elements repeated in the
RT bacterial chromosome."
RL J. Mol. Biol. 196:445-455(1987).
CC -1- FUNCTION: INVOLVED IN THE TRANSPOSITION OF THE INSERTION
CC SEQUENCE.
CC -1- SIMILARITY: BELONGS TO THE IS21/IS408/IS1162 FAMILY OF
CC TRANSPOSASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X05956; CAA29390.1; -
DR PIR: S03416; S03416.
DR InterPro: IPR001584; RVE.
DR Pfam: PF00665; RVE; 1.
KM Transposable element; Transposition; DNA-binding; DNA recombination.
SQ SEQUENCE 315 AA; 37544 MW; DB92FEC6777D1DA2 CRC64;
Query Match 46.7%; Score 42; DB 1; Length 315;
Best Local Similarity 50.0%; Pred. No. 8.2;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 PARRARYWVRCNP 15
Db 44 PARKTRKMKVRLKP 57
RESULT 5
ID ISTA_PSEAE STANDARD; PRT; 390 AA.
AC P15025;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transposase for insertion sequence element IS21.
GN IS7A.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Plasmid R68.45.
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89218951; Pubmed=2540414;
RA Reimann C., Moore R., Little S., Savioz A., Willetts N.S., Haas D.;
RT "Genetic structure, function and regulation of the transposable

RT element IS21.";
RL Mol. Gen. Genet. 215:416-424(1989).
RN [2]
RP REVISION TO 283.
RA Berger B.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE TRANSPORT OF THE INSERTION
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE IS21/IS408/IS1162 FAMILY OF
CC TRANSPOSASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X14793; CAA32898.2; -.
DR PIR: JY0012; BVEG1S.
DR InterPro: IPR001822; Recombinase.
DR InterPro: IPR001584; Rve.
DR Pfam: PF00665; rve; 1.
DR PROSITE: PS00397; RECOMBINASES_1; FALSE_NEG.
DR PROSITE: PS00398; RECOMBINASES_2; FALSE_NEG.
DR Plasmid; Transposable element; Transposition; DNA-binding;
KW DNA recombination.
FT DNA_BIND 20 DT
SQ SEQUENCE 390 AA: 45982 MW: 826B0037237C9070 CRC64;
H-T-H MOTIF (POTENTIAL).
Query Match 46.7%; Score 42; DB 1; Length 390;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 PARARYQWVRECKP 15
DB 44 PAKTRHMKVCLKP 57
IIII:IIII:IIII
ID EX12_HUMAN STANDARD; PRT; 394 AA.
AC Q9NWF9; 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DE 15-JUN-2002 (Rel. 41; Last annotation update)
DE Ethanolamine kinase-like protein EX12 (FLJ10761).
GN EX12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Nakamatsu A., Nakamura Y., Nagahari K.,
RA Masuko Y., Kanehori K.;
RT "NEO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE-21125782; PubMed-11044454;
RA Lykdis A., Wang J., Karim M.A., Jackowski S.;
RT "Overexpression of a mammalian ethanolamine-specific kinase
RT accelerates the CDP-ethanolamine pathway.";
RL J. Biol. Chem. 276:2174-2179(2001).
CC -!- SIMILARITY: BELONGS TO THE CHOLINE/ETHANOLAMINE KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AK001623; BAA91793.1; -.
DR InterPro: IPR002573; Choline_kinase.
DR Pfam: PF01633; Choline_kinase; 1.
KW Transferase; Kinase.
FT ACT_SITE 247 247 BY SIMILARITY.
SQ SEQUENCE 394 AA: 44871 MW: 5BD02C035622A81B CRC64;
Query Match 44.4%; Score 40; DB 1; Length 394;
Best Local Similarity 54.5%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 YPARARYQWV 11
DB 299 YPARETQLQWL 309
IIII:IIII:IIII
ID POLG_HCVA STANDARD; PRT; 3898 AA.
AC P19712;
DT 01-FEB-1991 (Rel. 17; Created)
DE 16-OCT-2001 (Rel. 40; Last sequence update)
DE 16-OCT-2001 (Rel. 40; Last annotation update)
DE Genome polyprotein.
OS Hog cholera virus (strain Alfort) (Swine fever virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_Taxid=11097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-89348014; PubMed-2763466;
RA Meyers G., Ruemenapf T., Thiel H.-J.;
RT "Molecular cloning and nucleotide sequence of the genome of hog
RT cholera virus.";
RL Virology 171:555-567(1989).
RN [2]
RP REVISION TO 2731.
RA Meyers G.;
RL Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: PESTIVIRUS P80 (P125) MAY BE A BIFUNCTIONAL PROTEIN
CC WITH HELICASE AND PROTEASE ACTIVITY.
CC -!- SUBCELLULAR LOCATION: THE GP51-GP54 PROTEIN IS ANCHORED TO THE
CC VIRAL ENVELOPE.
CC -!- SIMILARITY: TO BOVINE VIRAL DIARRHEA VIRUS GENOME POLYPROTEIN.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S31.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04358; AAA43844.2; -.
DR PIR: A34037; GNVVHC.
DR HSP: P27958; IAIY.
DR MEROPS: C53.001; -.
DR MEROPS: S31.001; -.
DR InterPro: IPR000280; CDvir_endptsep80.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR001568; RNase_T2.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; HCV_RDRP; 1.
DR PRINTS: PR00729; CDVENDOPTASE.

DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELICC; 1.
 DR PROSITE; PS00531; RNASE_T2.2; UNKNOWN_1.
 KW Polypeptide; Glycoprotein; Transmembrane; Hydrolase; Serine protease; Helicase.
 FT CHAIN 21 2267 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 7268 7500 GP4-GP48 (E2) (POTENTIAL).
 FT CHAIN 7501 7689 GP31 (E3) (POTENTIAL).
 FT CHAIN 7690 71060 GP55 (ENVELOPE PROTEIN E1).
 FT CHAIN 71611 72111 P80 (POTENTIAL).
 FT TRANSMEM 1032 1048 POTENTIAL.
 FT ACT_SITE 1658 1658 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1695 1695 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1752 1752 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 805 805 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 918 918 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 986 986 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1713 1713 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2134 2134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2217 2217 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2494 2494 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2787 2787 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2815 2815 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2891 2891 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3211 3211 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3316 3316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3689 3689 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3698 3698 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3794 3794 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 387 387 R -> A.
 FT VARIANT 3542 3542 T -> S.
 SQ SEQUENCE 3898 AA; 438570 MW; 2C1F17B8A359D0F6 CRC64;
 Query Match 43.9%; Score 39.5; DB 1; Length 3898;
 Best Local Similarly 36.4%; Pred. No. 2.5e+02;
 Matches 8; Conservative 3; Mismatches 4; Indels 7; Gaps 1;
 Oy 1 YPARRAR-----YQWYRKP 15
 Db 3469 YPEAKTRLATIKVMYKWKOKP 3490
 RESULT 8
 POLG_HCVB STANDARD; PRT; 3898 AA.
 AC P21530;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Genome polypeptide.
 OS Hog cholera virus (strain Brescia) (Swine fever virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Pestivirus.
 OX NCBI_Taxid=11098;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90281581; PubMed=2162104;
 RA Moormann R.J.M., Warmerdam P.A.M., van der Meer B., Schaper W.M.M.,

RA Wensvoort G., Hulst M.M.;
 RT "Molecular cloning and nucleotide sequence of hog cholera virus strain Brescia and mapping of the genomic region encoding envelope protein E1."
 RT Virology 177:184-198(1990).
 CC -I- FUNCTION: PESTIVIRUS P80 (P125) MAY BE A BIFUNCTIONAL PROTEIN WITH HELICASE AND PROTEASE ACTIVITY.
 CC -I- SUBCELLULAR LOCATION: THE GP51-GP54 PROTEIN IS ANCHORED TO THE VIRAL ENVELOPE.
 CC -I- SIMILARITY: TO BOVINE VIRAL DIARRHEA VIRUS GENOME POLYPEPTIDE.
 CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S31.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M31768; AAA3843.1; -.
 DR PIR; A35317; GNMVHB.
 DR HSSP; P27958; IALV.
 DR MEROPS; C53.001; -.
 DR MEROPS; S31.001; -.
 DR InterPro; IPR000280; Cpylr_endptseP80.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002166; HCV_RdRp.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR001568; RNase_T2.
 DR Pfam; PF00271; helicase_C.1.
 DR Pfam; PF00998; HCV_RdRp.1.
 DR PRINTS; PR00729; CDVENDOPTASE.
 DR SMART; SM00490; HELICC; 1.
 DR PROSITE; PS00531; RNASE_T2.2; UNKNOWN_1.
 KW Polypeptide; Glycoprotein; Transmembrane; Hydrolase; Serine protease; Helicase.
 FT CHAIN 21 2267 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 7268 7500 GP42 (E2) (POTENTIAL).
 FT CHAIN 7601 7689 GP31 (E3) (POTENTIAL).
 FT CHAIN 7690 71060 GP51-GP54 (ENVELOPE PROTEIN E1).
 FT CHAIN 71611 72111 P80 (POTENTIAL).
 FT TRANSMEM 1032 1048 POTENTIAL.
 FT ACT_SITE 1658 1658 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1695 1695 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1752 1752 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 805 805 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 918 918 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1713 1713 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2134 2134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2217 2217 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2419 2419 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2494 2494 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2787 2787 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2815 2815 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2891 2891 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3103 3103 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3211 3211 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3316 3316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3689 3689 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3698 3698 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 3794 3794 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 3898 AA; 438423 MW; EC6EB207A09D59PD CRC64;
 Query Match 43.9%; Score 39.5; DB 1; Length 3898;
 Best Local Similarity 36.4%; Pred. No. 2.5e+02;
 Matches 8; Conservative 3; Mismatches 4; Indels 7; Gaps 1;

QY 1 YPARRAR-----YQWRCKP 15
 11:11:11
 Db 3469 YPEAKTRLATIKYWKVKQKP 3490

RESULT 9
 Y4QJ_RHISN
 ID Y4QJ_RHISN STANDARD; PRT; 398 AA.
 AC P55631;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Putative transposase Y4QJ.
 GN Y4QJ.
 OS Rhizobium sp. (strain NGR234).
 OC Plasmid sym pNCR234.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97305956; PubMed-9163424;
 RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
 RA Perret X.;
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";
 RL Nature 387:394-401(1997).
 CC -1 SIMILARITY: TO P. SYRINGAE TRANSPOSASE FOR INSERTION SEQUENCE
 CC ELEMENT IS601.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000093; AAB92465.1; -
 KW Hypothetical protein; Transposable element; Transposition;
 KW DNA-binding; DNA recombination; Plasmid.
 SQ SEQUENCE 398 AA; 45804 MW; B77D198503597BC4 CRC64;

Query Match 43.3%; Score 39; DB 1; Length 398;
 Best Local Similarity 71.4%; Pred. No. 32;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 9 QWRCKP 15
 11:11:11
 Db 176 QWICKP 182

RESULT 10
 NTC2_RAT
 ID NTC2_RAT STANDARD; PRT; 2471 AA.
 AC Q90W30;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).
 GN NOTCH2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Brain;
 RX MEDLINE=93202015; PubMed-1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941(1992).
 RN [2]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed-11438922;
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-DSL signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -1 FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May play an essential role in postimplantation
 CC development, probably in some aspect of cell specification and/or
 CC differentiation (By similarity).
 CC -1 SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1 SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1 TISSUE SPECIFICITY: Highly expressed in the spleen and choroid
 CC plexus in the brain. Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within numerous
 CC cells throughout the CNS. It is more highly localized to
 CC ventricular germinal zones. Also found in the heart, liver and
 CC kidney.
 CC -1 DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.
 CC -1 PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase to yield
 CC trans-cold network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1 PTM: Phosphorylated (By similarity).
 CC -1 SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1 SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.
 CC -1 SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
 CC -1 SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M93661; AAK13558.1; -
 DR HSP: P00743; ICF.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxy1.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR001438; EGF-11.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR000800; Notch.
 DR Pfam: PF00008; EGF: 35.
 DR Pfam: PF00023; ank: 6.
 DR Pfam: PF00066; notch: 2.
 DR PRINTS: PR00010; EGFLOAD.


```

RT clustered together in an organization different from the same genes
RT in Acremonium chrysogenum and Penicillium chrysogenum."
RL Mol. Microbiol. 5:1125-1133(1991).
CC -1- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS
CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
CC INTERMEDIATES.
CC -1- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
CC CEPHALOSPORIN.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X57310; CAA40561.1; -
DR PIR: S18268; S18268.
DR HSSP: P14687; 1AMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR003880; Ppatine-attach.
DR InterPro: IPR000379; Ser-estr. site.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 3.
DR Pfam: PF00550; PP-binding; 3.
DR Pfam: PF00668; Condensation; 3.
DR Pfam: PF00975; Thioesterase; 1.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE: PS00455; AMP-BINDING; 1.
DR PROSITE: PS50075; ACP_DOMAIN; 3.
DR Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
KW Repeat; Phosphopantetheine.
FT REPEAT 401 861 DOMAIN 1 (ADIPATE-ACTIVATING).
FT REPEAT 1014 1937 DOMAIN 2 (CYSTEINE-ACTIVATING).
FT REPEAT 2079 2985 DOMAIN 3 (VALINE-ACTIVATING).
FT DOMAIN 788 857 ACTL CARRIER (ACP) 1.
FT DOMAIN 1864 1933 ACTL CARRIER (ACP) 2.
FT DOMAIN 2910 2981 ACTL CARRIER (ACP) 3.
FT BINDING 820 820 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 1896 1896 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2944 2944 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT ACT_SITE 3502 3502 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 3649 AA; 404079 MW; 6FD095704F858EBB CRC64;

Query Match 43.3%; Score 39; DB 1; Length 3649;
Best local Similarity 60.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 1 YPARRRARYQW 10
DB 931 YPARLRRLRFDM 940

RESULT 12
R15B_YEAST
ID R15B_YEAST STANDARD; PRT; 203 AA.
AC P54780;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L15-B (YL10) (L13) (RPL15R) (YP18).
GN RPL15B OR RPL10B OR YL10B OR YLR121C OR YMR564.03C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

```

```

OX NCBI_TaxID=4932;
RN
RN SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Lye G., Churcher C.M., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
RN
RN SEQUENCE OF 1-43.
RX MEDLINE=83048950; PubMed=6814480;
RA Otake E., Higo K.-I., Osawa S.;
RT "Isolation of seventeen proteins and amino-terminal amino acid
RT sequences of eight proteins from cytoplasmic ribosomes of yeast.";
RL Biochemistry 21:4545-4550(1982).
RN
RN SEQUENCE OF 1-8.
RX MEDLINE=92184799; PubMed=1544921;
RA Takakura H., Tsunasawa S., Miyagi M., Warner J.R.;
RT "NH2-terminal acetylation of ribosomal proteins of Saccharomyces
RT cerevisiae.";
RL J. Biol. Chem. 267:5442-5445(1992).
CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L15 IN YEAST.
CC -1- SIMILARITY: BELONGS TO THE L15 FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49273; CAA89270.1; -
DR SGD: S0004728; RPL15B.
DR InterPro: IPR000439; Ribosomal_L15e.
DR Pfam: PF00827; Ribosomal_L15e; 1.
DR PROSITE: PS01194; RIBOSOMAL_L15E; 1.
KW Ribosomal protein; Multigene family.
FT INIT_MET 0 0
FT CONFLICT 27 27 W -> G (IN REF. 2).
SQ SEQUENCE 203 AA; 24291 MW; AE472EB562931E2C CRC64;

Query Match 42.8%; Score 38.5; DB 1; Length 203;
Best local Similarity 58.3%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
OY 4 RRARYQWVRCKP 15
DB 143 RQARYNMI-CNP 153

RESULT 13
MML19_HUMAN
ID MML19_HUMAN STANDARD; PRT; 508 AA.
AC Q99542; Q99580; O15278; Q95606;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Matrix metalloproteinase-19 precursor (EC 3.4.24.-) (MMP-19) (Matrix
DE metalloproteinase RASI) (MMP-18).
GN MMP19 OR MMP18 OR RASI.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN
RN SEQUENCE FROM N.A. (TSOFORM 1).
RC TISSUE=Mammary gland;
RX MEDLINE=97079209; PubMed=8920941;
RA Cossins J., Dudgeon T.J., Catlin G., Gearing A.J.H., Clements J.M.;
RT "Identification of MMP-18, a putative novel human matrix
RT metalloproteinase.";
RL Biochem. Biophys. Res. Commun. 228:494-498(1996).
RN

```

RP SEQUENCE FROM N.A. (ISOFORM 1).
 RP TISSUE-LIVER:
 RX MEDLINE=97172501; PubMed-9020145;
 RA Pendas A.M., Kraeuper V.V., Puente X.S., Llano E., Mattei M.-G.,
 RA Apte S., Murphy G., Lopez-Otin C.;
 RT "Identification and characterization of a novel human matrix
 RT metalloproteinase with unique structural characteristics, chromosomal
 RT location and tissue distribution.";
 RT J. Biol. Chem. 272:4281-4286(1997).
 RN [13]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RP TISSUE-Rheumatoid arthritis synovial fluid;
 RC MEDLINE=97376320; PubMed-99232430;
 RX Kolb C., Mauch S., Peter H.-H., Kraewinkel U., Sedlacek R.,
 RA "The matrix metalloproteinase RAS1-1 is expressed in synovial blood
 RT vessels of a rheumatoid arthritis patient.";
 RT Immunol. Lett. 57:83-88(1997).
 RN [4]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RP TISSUE-Rheumatoid arthritis synovial fluid;
 RC MEDLINE=98223959; PubMed-9562866;
 RX Sedlacek R., Mauch S., Kolb B., Schaezelzlein C., Eibel H., Peter H.-H.,
 RA Schmitt J., Kraewinkel U.;
 RT "Matrix metalloproteinase MMP-19 (RAS1-1) is expressed on surface of
 RT activated peripheral blood mononuclear cells and is detected as an
 RT autoantigen in rheumatoid arthritis.";
 RT Immunobiology 198:408-423(1998).
 RN [5]
 RN SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RP Much S., Sedlacek R., Kraewinkel U., Schaezelzlein C.;
 RL Submitted (OCT-1995) to the EMBL/Genbank/DBD databases.
 RN [6]
 RN SEQUENCE OF N-TERMINUS, FUNCTION, AND MUTAGENESIS.
 RP MEDLINE=20270177; PubMed-10809722;
 RX Stracke J.O., Hutton M., Stewart M., Pendas A.M., Smith B.,
 RA Lopez-Otin C., Murphy G., Kraeuper V.;
 RT "Biochemical characterization of the catalytic domain of human matrix
 RT metalloproteinase 19. Evidence for a role as a potent basement
 RT membrane degrading enzyme.";
 RT J. Biol. Chem. 275:14809-14816(2000).
 RN [7]
 RN FUNCTION.
 RP MEDLINE=20382723; PubMed-10922468;
 RX Stracke J.O., Fosang A.J., Last K., Mercuri F.A., Pendas A.M.,
 RA Llano E., Perris R., Di Cesare P.E., Murphy G., Kraeuper V.;
 RT "Matrix metalloproteinases 19 and 20 cleave aggrecan and cartilage
 RT oligomeric matrix protein (COMP).";
 RT FEBS Lett. 478:52-56(2000)
 CC -1- FUNCTION: ENDOPEPTIDASE THAT DEGRADES VARIOUS COMPONENTS OF THE
 CC EXTRACELLULAR MATRIX, SUCH AS AGGRECAN AND CARTILAGE OLIGOMERIC
 CC MATRIX PROTEIN (COMP), DURING DEVELOPMENT, HAEMOSTASIS AND
 CC PATHOLOGICAL CONDITIONS (ARTHRITIC DISEASE). MAY ALSO PLAY A ROLE
 CC IN NEOVASCULARIZATION OR ANGIOGENESIS. HYDROLYZES COLLAGEN TYPE
 CC IV, LAMININ, NIDOGEN, NASCIN-C ISOFORM, FIBRONECTIN, AND TYPE I
 CC GELATIN.
 CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 360-Ser-1-Phe-361
 CC site.
 CC -1- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY (BY SIMILARITY).
 CC -1- ENZYME REGULATION: STRONGLY INHIBITED BY TIMP-2, TIMP-3 AND TIMP-
 CC -4, WHILE TIMP-1 IS LESS EFFICIENT.
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1/RAS1-1 (SHOWN HERE),
 CC 2/RAS1-6 AND 3/RAS1-9; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MAMMARY GLAND, PLACENTA, LUNG,
 CC PANCREAS, OVARY, SMALL INTESTINE, SPLEEN, THYRUS, PROSTATE, TESTIS
 CC COLON, HEART AND BLOOD VESSEL WALLS. NOT DETECTED IN BRAIN AND
 CC PERIPHERAL BLOOD LEUKOCYTES. ALSO EXPRESSED IN THE SYNOVIAL FLUID
 CC OF NORMAL AND RHEUMATOID PATIENTS.
 CC -1- PTM: AUTOACTIVATES AT THE 97-TYR-98 SITE.
 CC -1- DISEASE: MAY PLAY A ROLE IN PATHOLOGICAL PROCESSES PARTICIPATING
 CC IN RHEUMATOID ARTHRITIS (RA)-ASSOCIATED JOINT TISSUE DESTRUCTION.
 CC AUTOANTIGEN ANTI-MMP19 IS FREQUENT IN RA PATIENTS.

```

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A.
CC -1- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC -1- CAUTION: REF.5 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y08622; CAA69913.1; -
CC DR EMBL; X92521; CAA63299.1; -
CC DR EMBL; U37791; AAC51521.1; -
CC DR EMBL; U38321; AAB63008.1; -
CC DR EMBL; U38431; AAC99995.1; ALT_FRAME.
CC DR EMBL; U38322; AAB63009.1; -
CC DR HSSP; P03956; IHFC.
CC DR MEROPS; M10.021; -
CC DR Genew; HGNC:7165; MMP19.
CC DR MIM; 601807; -
CC DR InterPro; IPR000585; Hemopexin.
CC DR InterPro; IPR001818; Matrxin.
CC DR InterPro; IPR000130; Zn_mtpetpde.
CC DR Pfam; PF00045; hemopexin; 4.
CC DR Pfam; PF00413; peptidase_M10; 1.
CC DR PRINTS; PRO0138; MATRIXIN.
CC DR SMART; SM00230; HX; 4.
CC DR SMART; SM00325; Zmc; 1.
CC DR PROSITE; PS00546; CYSTEINE_SWITCH; FALSE_NEG.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
CC DR PROSITE; PS00024; HEMOEXIN; FALSE_NEG.
CC DR HydroLase; Metalloprotease; zinc; Calcium; Zymogen; Signal;
CC KW Glycoprotein; Extracellular matrix; Collagen degradation;
CC KW Alternative splicing.
CC KM SIGNAL 1 18 POTENTIAL.
CC FT PROPEP 19 97
CC FT CHAIN 98 508 MATRIX METALLOPROTEINASE-19.
CC FT DOMAIN 286 475 HEMOEXIN-LIKE.
CC FT SITE 85 85 CYSTEINE SWITCH (POTENTIAL).
CC FT METAL 212 212 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 213 213 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 216 216 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 222 222 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT DOMAIN 265 271 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT DISULFD 289 472 POLY-GLU.
CC FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT VASPPIC 1 286 MISSING (IN ISOFORM 3).
CC FT VASPPIC 287 298 DPCSELDAMML -> MGVTWDFSMNG (IN ISOFORM
CC FT VASPPIC 1 79 MISSING (IN ISOFORM 2).
CC FT MUTAGEN 88 88 E->P. REDUCED THE AUTOLYSIS RATE.
CC FT MUTAGEN 90 90 P->Y. REDUCED THE AUTOLYSIS RATE.
CC FT CONFLICT 376 376 V -> S (IN REF. 1).
CC SQ SEQUENCE 508 AA; 57357 MM; BA480549AA9A8972 CRC64;
CC
CC Query Match 42.8%; Score 38.5; DB 1; Length 508;
CC Best Local Similarity 40.0%; Pred. No. 49;
CC Matches 6; Conservative 3; Mismatches 5; Indels 1; Gaps 1;
CC
CC Oy 1 YPARARYQWVCKP 15
CC ||| : ||: ||
CC Db 461 YP-RNISHMMHCRP 474
CC
CC RESULT 14
CC YMBW_YEAST STANDARD: PRT: 150 AA.
CC AC Q03579;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

```

DT 01-NOV-1997 (rel. 35, Last annotation update)
DE Hypothetical 17.2 kDa protein in PRC1-ADP4 intergenic region.
GN YMR298W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80836; CAA56807.1; -
DR SGD; S0004913; YMR298W.
KM Hypothetical protein.
SQ SEQUENCE 150 AA; 17207 MW; 75DBA3525C3065C CRC64;

Query Match
Best Local Similarity 42.2%; Score 38; DB 1; Length 150;
Matches 5; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 RARYOWRCKP 15
   1 1 1 1 1
Db 45 RINWEFHCTP 55

RESULT 15
RNHL_YEAST
ID RNHL_YEAST STANDARD; PRT; 307 AA.
AC P53942;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Ribonuclease H1 35 large subunit (EC 3.1.26.-) (RNase H1 large
DE subunit) (RNase H(35)).
GN RNH35 OR YNL072W OR NZ369.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=96267764; PubMed=8701611;
RA Poehlmann R., Philippsen P.;
RT "Sequencing a cosmid clone of Saccharomyces cerevisiae chromosome XIV
RT reveals 12 new open reading frames (ORFs) and an ancient duplication
RT of six ORFs."
RL Yeast 12:391-402(1996).
RN (12)
RP CHARACTERIZATION.
RX MEDLINE=98122664; PubMed=9462832;
RA Frank P., Braunschöfer-Reiter C., Wintersberger U.;
RT "Yeast RNase H(35) is the counterpart of the mammalian RNase H1, and
RT is evolutionarily related to prokaryotic RNase H1."
RL FEBS Lett. 421:23-26(1998).
CC -1- FUNCTION: DEGRADATES THE RIBONUCLEOTIDE MOIETY ON RNA-DNA HYBRID
CC MOLECULES. PARTICIPATES IN DNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomonoester.
CC -1- SIMILARITY: BELONGS TO THE RNASE H11 FAMILY. EUKARYOTIC
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X86470; CAA60188.1; -
DR DR EMBL; 271348; CAA95946.1; -
DR HSSP; O57599; IERF.
DR SGD; S0005016; YNL072W.
DR InterPro; IPR001352; RNase_H11/H111.
DR InterPro; IPR004649; Rnh11.
DR Pfam; PF01351; RNase_H11; 1.
DR TIGRPFAM; TIGR00729; Rnh11; 1.
KM Hydrolyase; Nuclease; Endonuclease.
FT ACT_SITE 39 39 BY SIMILARITY.
FT ACT_SITE 155 155 BY SIMILARITY.
FT ACT_SITE 183 183 BY SIMILARITY.
SQ SEQUENCE 307 AA; 34875 MW; 60B1005F674ECC88 CRC64;

Query Match
Best Local Similarity 42.2%; Score 38; DB 1; Length 307;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 YPARRARIOWRCK 14
   1 1 1 1 1
Db 239 WPAWVRFMSWTCQ 252

```

Search completed: April 8, 2003, 11:24:40
 Job time : 13.6379 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: Apr11 8, 2003, 11:18:53 ; Search time 23.7931 seconds
(without alignments)
60.606 Million cell updates/sec

Title: US-09-496-391-9

Perfect score: 90

Sequence: 1 YPARRRARYQWVRCKP 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	100.0	152	2 J00791	proteoglycan 17K c
2	90	100.0	179	0 GZRT0	secretory granule
3	77	85.6	158	2 A28058	secretory granule
4	45	50.0	91	2 T06502	glycine-rich prote
5	44	48.9	111	2 T14306	hypothetical prote
6	44	48.9	328	2 T41074	hypothetical prote
7	43	47.8	1708	2 F69189	protoporphyrin IX
8	42.5	47.2	7576	2 T17438	PK506 polyketide s
9	42	46.7	195	2 H6532	hypothetical prote
10	42	46.7	315	2 S03416	probable transpos
11	42	46.7	390	1 BVEG1S	ista protein - Esc
12	42	46.7	390	2 T43607	probable transpos
13	41	45.6	204	2 E83494	hypothetical prote
14	41	45.6	398	2 T25719	hypothetical prote
15	41	45.6	458	2 C72581	hypothetical prote
16	40	44.4	896	2 AE1785	the two components
17	40	44.4	896	2 AE1785	the two components
18	39.5	43.9	3898	1 GNMVHC	genome polypotein
19	39.5	43.9	3898	1 GNMVHC	genome polypotein
20	39.5	43.9	3898	2 S57437	genome polypotein
21	39.5	43.9	3898	2 S58295	genome polypotein
22	39	43.3	209	2 T46140	hypothetical prote
23	39	43.3	304	2 T29421	hypothetical prote
24	39	43.3	369	2 T29836	hypothetical prote
25	39	43.3	448	2 T15188	hypothetical prote
26	39	43.3	719	2 A67292	primosomal prote
27	39	43.3	771	2 A63348	probable aldehyde
28	39	43.3	2471	2 A49128	cell-fate determin
29	39	43.3	3649	1 S18268	delta-(L-alpha-aml

30	38.5	42.8	204	2 S54490	ribosomal protein
31	38.5	42.8	343	2 T29547	hypothetical prote
32	38.5	42.8	425	2 D75164	hypothetical prote
33	38.5	42.8	428	2 G71177	hypothetical prote
34	38.5	42.8	508	2 J05082	matrix metallopro
35	38.5	42.8	925	2 T02811	DNA excision/repai
36	38	42.2	150	2 S47459	probable membrane
37	38	42.2	231	2 H85138	hypothetical prote
38	38	42.2	292	2 T02320	hypothetical prote
39	38	42.2	307	2 S53908	hypothetical prote
40	38	42.2	330	2 A63650	sucnioglycan bios
41	38	42.2	495	2 A54980	easily shocked pro
42	38	42.2	525	2 E75374	conserved hypothet
43	37.5	41.7	204	2 S48502	ribosomal protein
44	37.5	41.7	535	2 AFO103	probable sulfatase
45	37	41.1	123	2 S72927	hypothetical prote

ALIGNMENTS

RESULT 1
J00791
proteoglycan 17K core protein precursor - mouse
N:Alternate names: secretory granule proteoglycan core protein; serglycin
C:Species: Mus musculus (house mouse)
C:Date: 12-Feb-1993 #sequence revision 12-Feb-1993 #text_change 13-Aug-1999
C:Accession: J00791; A34399; S06127; A32779
R:Angerth, T.; Huang, R.; Aveskog, M.; Pettersson, I.; Kjellen, L.; Hellman, L.
Gene 93, 235-240, 1990
A>Title: Cloning and structural analysis of a gene encoding a mouse mastocytoma prote

A:Reference number: J00791; MUID:91033033; PMID:2121613

A:Accession: J00791

A:Molecule type: DNA

A:Residues: 1-152 <ANG>

A:Cross-references: GB:M27393; GB:J05044; NID:g200458; PIDN:AAA39965.1; PID:g387517

A:Experimental source: mast cell

A:Title: Primary structure of a mouse mastocytoma proteoglycan core protein.

A:Reference number: S06127; MUID:90104216; PMID:2532501

A:Accession: S06127

A:Molecule type: mRNA

A:Residues: 1-152 <KUE>

A:Cross-references: EMBL:X16133; NID:g54085; PIDN:CAA34259.1; PID:g54086

A:Avraham, S.; Stevens, R.L.; Nicodemus, C.F.; Gartner, M.C.; Austen, K.F.; Weis, J.H

Proc. Natl. Acad. Sci. U.S.A. 86, 3763-3767, 1989

A:Title: Molecular cloning of a cDNA that encodes the peptide core of a mouse mast ce

A:Reference number: A32779; MUID:89264503; PMID:2726751

A:Accession: A32779

A:Molecule type: mRNA

A:Residues: 1-152 <AV2>

A:Cross-references: GB:J04549; NID:g200973; PIDN:AAA40111.1; PID:g200974

C:Genetics:

A:Gene: SGC

A:Introns: 25/1; 73/2

C:Superfamily: proteoglycan 10K core protein

F:1-25/Domain: signal sequence #status predicted <STG>

F:26-152/Product: proteoglycan 17K core protein #status predicted <MAY>

Query Match 100.0%; Score 90; DB 2; Length 152;

Best Local Similarity 100.0%; Pred. No. 4.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPARARYQWVRCCKP 15
DB 26 YPARARYQWVRCCKP 40

RESULT 2
GZRT0
secretory granule proteoglycan core protein precursor - rat
N:Alternate names: 19PG core protein; chondroitin sulfate proteoglycan core protein; RBL
C:Species: Rattus norvegicus (Norway rat)
C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 22-Jun-1999
C:Accession: A25644; A28059; A60653
R:Boudon, M.A.; Shiga, M.; Rostalet, E.
J. Biol. Chem. 261, 12534-12537, 1986
A>Title: Identification from cDNA of the precursor form of a chondroitin sulfate proteog
A:Reference number: A25644; MUID:86304425; PMID:2427521
A:Accession: A25644
A:Molecule type: mRNA
A:Residues: 1-179 <BOU>
A:Cross-references: GB:K02934; GB:M14282; NID:g207039; PIDN:AAA42171.1; PID:g207040
A:Experimental source: yolk sac carcinoma
R:Avraham, S.; Stevens, R.L.; Gartner, M.C.; Austen, K.F.; Lalley, P.A.; Wels, J.H.
J. Biol. Chem. 263, 7292-7296, 1988
A>Title: Isolation of a cDNA that encodes the peptide core of the secretory granule prot
A:Reference number: A28059; MUID:88213412; PMID:3366780
A:Accession: A28059
A:Molecule type: mRNA
A:Residues: 1-179 <AVR>
A:Cross-references: EMBL:J03224; NID:g206108; PIDN:AAA41837.1; PID:g206109
R:Giorda, R.; Chambers, W.H.; Dahl, C.A.; Trucco, M.
Nat. Immun. Cell Growth Regul. 9, 91-102, 1990
A>Title: Isolation and characterization of a cDNA that encodes the core protein of the c
A:Reference number: A60653; MUID:90279692; PMID:2352541
A:Accession: A60653
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-179 <GIO>
C:Superfamily: proteoglycan 10K core protein
C:Keywords: chondroitin sulfate proteoglycan; dermatan sulfate; extracellular matrix; fi
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-75/Domain: activation peptide #status predicted <APR>
F:76-179/Product: proteoglycan 10K core protein #status predicted <MPR>
F:90-138/Region: chondroitin sulfate attachment (S-G repeats)

Query Match 100.0%; Score 90; DB 1; Length 179;
Best Local Similarity 100.0%; Pred. No. 4; 8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YPARARYQWVRCCKP 15
DB 27 YPARARYQWVRCCKP 41

RESULT 3
A28058
secretory granule proteoglycan core protein precursor - human
N:Alternate names: HL-60 cell proteoglycan peptide core; platelet proteoglycan
C:Species: Homo sapiens (man)
C>Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 13-Aug-1999
C:Accession: A35183; A28058; S09610; S01126; S01819
R:Nicodemus, C.F.; Avraham, S.; Austen, K.F.; Purdy, S.; Jablonski, J.; Stevens, R.L.
J. Biol. Chem. 265, 5889-5896, 1990
A>Title: Characterization of the human gene that encodes the peptide core of secretory g
A:Reference number: A35183; MUID:90202841; PMID:2180935
A:Accession: A35183
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-158 <NIC>
A:Cross-references: GB:J05824
R:Stevens, R.L.; Avraham, S.; Gartner, M.C.; Bruns, G.A.P.; Austen, K.F.; Wels, J.H.
J. Biol. Chem. 263, 7287-7291, 1988
A>Title: Isolation and characterization of a cDNA that encodes the peptide core of the s

A:Reference number: A28058; MUID:88213411; PMID:2835370
A:Accession: A28058
A:Molecule type: mRNA
A:Residues: 1-158 <STE>
A:Cross-references: GB:J03223; NID:g190419; PIDN:AAA60179.1; PID:g190420
R:Stellrecht, C.M.; Saunders, G.F.
Nucleic Acids Res. 17, 7523, 1989
A>Title: Nucleotide sequence of a cDNA encoding a hemopoietic proteoglycan core prote
A:Reference number: S09610; MUID:90016819; PMID:2798108
A:Accession: S09610
A:Molecule type: mRNA
A:Residues: 1-158 <ST2>
A:Cross-references: EMBL:X17042; NID:g32432; PIDN:CAA34900.1; PID:g32433
R:Alillet, P.M.; Perin, J.P.; Maillet, P.; Bonnet, F.; Rosa, J.P.; Jolles, P.
FEBS Lett. 236, 123-126, 1988
A>Title: Complete amino acid sequence of a human platelet proteoglycan.
A:Reference number: S01126; MUID:88296856; PMID:3402609
A:Accession: S01126
A:Molecule type: mRNA
A:Residues: 28-138, 'S', 140-158 <ALL>
A:Cross-references: EMBL:X12765
A>Note: Part of this sequence was confirmed by protein sequencing
R:Perin, J.P.; Bonnet, F.; Maillet, P.; Jolles, P.
Biochem. J. 255, 1007-1013, 1988
A>Title: Characterization and N-terminal sequence of human platelet proteoglycan.
A:Reference number: S01819; MUID:89104992; PMID:3214420
A:Accession: S01819
A:Molecule type: protein
A:Residues: 28-48, 'X', 50-93 <PER>
C:Genetics:
A:Gene: GDB:PRG1; PRG
A:Cross-references: GDB:120312; OMIM:177040
A:Map position: 10q22.1-10q22.1
C:Superfamily: proteoglycan 10K core protein
C:Keywords: chondroitin sulfate proteoglycan
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-158/Product: secretory granule proteoglycan core protein #status predicted <MAT>

Query Match 85.6%; Score 77; DB 2; Length 158;
Best Local Similarity 80.0%; Pred. No. 56-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPARARYQWVRCCKP 15
DB 28 YPTQARYQWVRCNP 42

RESULT 4
T06502
hypothetical protein 91 - garden pea chloroplast
C:Species: chloroplast Pisum sativum (garden pea)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: A23041; T06502
R:Rasmussen, O.F.; Stummann, B.M.; Henningsen, K.W.
Nucleic Acids Res. 12, 9143-9153, 1984
A>Title: Nucleotide sequence of a 1.1 kb fragment of the pea chloroplast genome conta
A:Reference number: A23041; MUID:85087925; PMID:6096819
A:Accession: A23041
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-91 <RMS>
A:Cross-references: EMBL:X01676; NID:g14204; PIDN:CAA25831.1; PID:g14205
C:Genetics:
A:Genome: chloroplast
C:Keywords: chloroplast

Query Match 50.0%; Score 45; DB 2; Length 91;
Best Local Similarity 41.7%; Pred. No. 3.7;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 PARARYQWVRC 13
I:::||||:|

hypothetical protein F14J22.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H96532
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luro, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; M0ID:21016719; PMID:11130712
A:Accession: H96532
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-195 <STO>
A:Cross-references: GB:AE005173; NID:g10120423; PIDN:AA613048.1; GSPDB:GN00141
C:Genetics:
A:Gene: F14J22.14
A:Map position: 1

Query Match 46.7%; Score 42; DB 2; Length 195;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 7 RYQWVRCRP 15
Db 187 RYQWVSLKP 195

RESULT 10
S03416
probable transposase - Shigella sonnei insertion sequence IS640
C:Species: Shigella sonnei
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 26-Aug-1999
C:Accession: S03416
R:Matsumi, S.; Ohtsubo, H.; Maeda, Y.; Ohtsubo, E.
J. Mol. Biol. 196, 445-455, 1987
A:Title: Isolation and characterization of IS elements repeated in the bacterial chromos
A:Reference number: S03411; M0ID:88062685; PMID:2824781
A:Accession: S03416
A:Molecule type: DNA
A:Residues: 1-315 <MAT>
A:Cross-references: EMBL:X05956; NID:g47543; PIDN:CAA29390.1; PID:g47544
C:Genetics:
A:Mobile element: insertion sequence IS640
C:Superfamily: Ista protein
C:Keywords: DNA binding

Query Match 46.7%; Score 42; DB 2; Length 315;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 PARRARYQWVRCRP 15
Db 44 PARKTRHKWVKLP 57

RESULT 11
BVCIS
Ista protein - Escherichia coli plasmid R68.45 insertion sequence IS21
C:Species: Escherichia coli
C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jul-1999
C:Accession: JY0012
R:Reilmann, C.; Moore, R.; Little, S.; Savioz, A.; Willetts, N.S.; Haas, D.
Mol. Gen. Genet. 215, 416-424, 1989
A:Title: Genetic structure, function and regulation of the transposable element IS21.
A:Reference number: A93128; M0ID:89218951; PMID:2540414
A:Accession: JY0012
A:Molecule type: DNA

A:Residues: 1-390 <RET>
A:Cross-references: GB:X14793; NID:g41826; PIDN:CAA32898.1; PID:g41827
A:Note: The authors translated the codon CCG for residue 283 as Leu
A:Note: It is uncertain whether Met-1 or Met-9 is the initiator
C:Comment: The ista protein, one of two proteins expressed only when there is a tandem
uch as plasmid R68.45, which is derived from tandem duplication of IS21 on plasmid R6
C:Genetics:
A:Gene: ista
A:Genome: plasmid R68.45
A:Mobile element: insertion sequence IS21
C:Superfamily: Ista protein

Query Match 46.7%; Score 42; DB 1; Length 390;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 PARRARYQWVRCRP 15
Db 44 PARKTRHKWVKLP 57

RESULT 12
T43607
probable transposase - Yersinia pestis plasmid pCD1
C:Species: Yersinia pestis
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T43607
R:Hu, P.; Elliott, J.; McCreedy, P.; Skowronski, E.; Garnes, J.; Kobayashi, A.; Bruba
J. Bacteriol. 180, 5192-5202, 1998
A:Title: Structural organization of virulence-associated plasmids of Yersinia pestis.
A:Reference number: Z22578; M0ID:98422474; PMID:9748454
A:Accession: T43607
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-390 <RUP>
A:Cross-references: EMBL:AF053946; NID:g2996222; PIDN:AA62589.1; PID:g2996266
A:Experimental source: strain KIM
C:Genetics:
A:Genome: plasmid pCD1
C:Superfamily: Ista protein

Query Match 46.7%; Score 42; DB 2; Length 390;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 PARRARYQWVRCRP 15
Db 44 PARKTRHKWVKLP 57

RESULT 13
E83494
hypothetical protein PA1219 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83494
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; M0ID:20437337; PMID:10984043
A:Accession: E83494
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <STO>
A:Cross-references: GB:AE004551; GB:AE004091; NID:g9947135; PIDN:AA604608.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1219

Query Match 45.6%; Score 41; DB 2; Length 204;
Best Local Similarity 63.8%; Pred. No. 32;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 4 RRARYQWVRC 14
 ||| |||
 Db 143 RLARASWVRCR 153

RESULT 14

T25719
 hypothetical protein F21F3.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T25719
 R:Geisels, C.; Kramer, J.; Elliott, G.
 submitted to the EMBL data library, February 1997
 A:Description: The sequence of C. elegans cosmid F21F3.
 A:Reference number: 220075
 A:Accession: T25719
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-398 <GEI>
 A:Cross-references: EMBL:U08175; PIDN:ABA42281.1; GSPDB:GN00019; CESP:F21F3.4
 A:Experimental source: strain Bristol N2; clone F21F3
 C:Genetics:
 A:Gene: CESP:F21F3.4
 A:Map position: 1
 A:Introns: 22/3; 43/2; 103/3; 231/3; 311/1
 C:Superfamily: Caenorhabditis elegans hypothetical protein F21F3.4

Query Match 45.6%; Score 41; DB 2; Length 398;
 Best Local Similarity 53.8%; Pred. No. 56;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 ARARYQWVRC 15
 :|||: |||
 Db 186 SRIRPRWVSC 198

RESULT 15

C72591
 hypothetical protein APE1199 - Aeropyrum pernix (strain KI)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: C72591
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6: 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
 A:Reference number: A72450; MUID:99310339; PMID:10382966
 A:Accession: C72591
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-458 <KAW>
 A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA80185.1; PID:95104871
 A:Experimental source: strain KI
 C:Genetics:
 A:Gene: APE1199

Query Match 45.6%; Score 41; DB 2; Length 458;
 Best Local Similarity 61.5%; Pred. No. 63;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 YPARARYQWVRC 13
 ||||| |
 Db 111 YPARRVRCWVSC 123

Search completed: April 8, 2003, 11:29:57
 Job time : 26.7931 secs

THIS PAGE BLANK (USPTO)


```
; ORGANISM: Homo sapiens
US-09-796-692-1379
Query Match
Best local Similarity 85.6%; Score 77; DB 9; Length 69;
Pred. No. 4,7e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YPARARYQWVRCKP 15
   ||:|||||||
Db 25 YPTORARYQWVRCKP 39

RESULT 2
US-09-796-692-2104
; Sequence 2104, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2104
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2104
Query Match
Best local Similarity 85.6%; Score 77; DB 9; Length 69;
Pred. No. 4,7e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YPARARYQWVRCKP 15
   ||:|||||||
Db 25 YPTORARYQWVRCKP 39

RESULT 3
US-09-796-692-2355
; Sequence 2355, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2355
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2355
Query Match
Best local Similarity 85.6%; Score 77; DB 9; Length 69;
Pred. No. 4,7e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YPARARYQWVRCKP 15
   ||:|||||||
Db 25 YPTORARYQWVRCKP 39

RESULT 4
US-09-796-692-1346
; Sequence 1346, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
```



```
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1346
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1346
```

```
Query Match      85.6%; Score 77; DB 9; Length 76;
Best Local Similarity 80.0%; Pred. No. 5.1e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 YPARARYQWRCMP 15
        ||:|||||||
Db      32 YPTQARYQWRCMP 46
```

```
RESULT 5
US-09-796-692-1835
; Sequence 1835, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1835
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1835
```

```
Query Match      85.6%; Score 77; DB 9; Length 76;
Best Local Similarity 80.0%; Pred. No. 5.1e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 YPARARYQWRCMP 15
        ||:|||||||
Db      32 YPTQARYQWRCMP 46
```

```
RESULT 6
US-09-796-692-2088
; Sequence 2088, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2088
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2088
```

```
Query Match      85.6%; Score 77; DB 9; Length 76;
Best Local Similarity 80.0%; Pred. No. 5.1e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      1 YPARARYQWRCMP 15
        ||:|||||||
Db      32 YPTQARYQWRCMP 46
```

```
RESULT 7
US-09-796-692-2318
; Sequence 2318, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
```

```
FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2318
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2318

Query Match      85.6%; Score 77; DB 9; Length 76;
Best Local Similarity 80.0%; Pred. No. 5.1e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 YPARARYOWRCNP 15
DB 32 YPTORARYOWRCNP 46

RESULT 8
US-09-796-692-1360
; Sequence 1360, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2329
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2329

Query Match      85.6%; Score 77; DB 9; Length 123;
Best Local Similarity 80.0%; Pred. No. 7.9e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```



```

; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1364
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(148)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-796-692-1364
```

```

Query Match      85.6%; Score 77; DB 9; Length 148;
Best Local Similarity 80.0%; Pred. No. 9.4e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 YPARRRARYQWRCRP 15
      11 :|||||
DB      32 YPTQRRARYQWRCNP 46
```

```

RESULT 13
US-09-796-692-2357
; Sequence 2357, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077, 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
```

```

; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2357
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(156)
; OTHER INFORMATION: Xaa = Any amino acid
US-09-796-692-2357
```

```

Query Match      85.6%; Score 77; DB 9; Length 156;
Best Local Similarity 80.0%; Pred. No. 9.9e-05;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 YPARRRARYQWRCRP 15
      11 :|||||
DB      26 YPTQRRARYQWRCNP 40
```

```

RESULT 14
US-09-796-692-1356
; Sequence 1356, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; FILE REFERENCE: 2077, 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1356
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variant
; LOCATION: (1)...(158)
```

; OTHER INFORMATION: Xaa = Any amino acid
US-09-796-692-1356

Query Match 85.6%; Score 77; DB 9; Length 158;
Best Local Similarity 80.0%; Pred. No. 0.0001;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPARRARYQWVRCNP 15
||:|||||||
Db 28 YPTQRRARYQWVRCNP 42

RESULT 15
US-09-796-692-2325

; Sequence 2325, Application US/09796692
; Publication No. US20020198362A1

; GENERAL INFORMATION:

; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077 001200

; CURRENT APPLICATION NUMBER: US/09/796,692

; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 60/200,303

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,999

; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 60/202,084

; PRIOR FILING DATE: 2000-05-04

; PRIOR APPLICATION NUMBER: 60/206,201

; PRIOR FILING DATE: 2000-05-22

; PRIOR APPLICATION NUMBER: 60/218,950

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: 60/222,903

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: 60/223,416

; PRIOR FILING DATE: 2000-08-04

; PRIOR APPLICATION NUMBER: 60/223,378

; PRIOR FILING DATE: 2000-08-07

; SOFTWARE: FASTSEQ for Windows Version 3.0

; SEQ ID NO 2325

; LENGTH: 158

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: variant

; LOCATION: (1)...(158)

; OTHER INFORMATION: Xaa = Any amino acid

US-09-796-692-2325

Query Match 85.6%; Score 77; DB 9; Length 158;
Best Local Similarity 80.0%; Pred. No. 0.0001;
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YPARRARYQWVRCNP 15
||:|||||||
Db 28 YPTQRRARYQWVRCNP 42

Search completed: April 8, 2003, 11:52:44
Job time : 21.6897 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:19:38 ; Search time 19.6552 Seconds
(Without alignments)
22.454 Million cell updates/sec

Title: US-09-496-391-9

Perfect score: 90

Sequence: 1 YPARRARYQWVRCRP 15

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilltest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	100.0	152	1	US-07-906-871-14
2	77	85.6	158	1	US-07-906-871-10
3	77	85.6	158	1	US-07-906-871-12
4	61	67.8	16	1	US-07-906-871-4
5	40	44.4	108	4	US-08-591-632-69
6	40	44.4	108	4	US-09-611-451-69
7	39.5	43.9	3898	2	US-08-876-991-2
8	39.5	43.9	3898	2	US-09-059-853-2
9	39.5	43.9	3898	4	US-08-750-717-2
10	39	43.3	229	4	US-09-267-177-26
11	38.5	42.8	508	4	US-09-391-104-18
12	38	42.2	2353	4	US-08-984-709A-50
13	37	41.1	861	1	US-08-484-105-18
14	37	41.1	861	1	US-08-484-106-18
15	36.5	40.6	593	4	US-09-000-145-4
16	36	40.0	41	1	US-08-111-939-27
17	36	40.0	51	4	US-09-336-536-71
18	36	40.0	249	4	US-09-336-536-42
19	36	40.0	365	4	US-09-336-536-40
20	36	40.0	394	4	US-09-336-536-39
21	36	40.0	521	2	US-08-682-847-4
22	35.5	39.4	24	4	US-09-082-358B-21
23	35.5	39.4	542	1	US-07-814-964-13
24	35.5	39.4	542	1	US-08-258-442-13
25	35.5	39.4	542	1	US-08-328-809-8
26	35.5	39.4	542	5	PCT-US92-1107-13
27	35.5	39.4	4302	3	US-08-658-136-5

28	35.5	39.4	4302	4	US-09-052-469-8	Sequence 8, Appli
29	35.5	39.4	4303	2	US-08-460-751-2	Sequence 2, Appli
30	35.5	39.4	4339	4	US-09-052-469-6	Sequence 6, Appli
31	35	38.9	20	1	US-08-023-760A-2	Sequence 2, Appli
32	35	38.9	247	4	US-08-961-083-170	Sequence 170, App
33	35	38.9	274	1	US-08-350-325A-7	Sequence 7, Appli
34	35	38.9	274	5	PCT-US94-03856-7	Sequence 3, Appli
35	35	38.9	467	1	US-08-459-100A-3	Sequence 3, Appli
36	35	38.9	531	2	US-08-975-114A-4	Sequence 4, Appli
37	35	38.9	531	4	US-08-849-281A-4	Sequence 4, Appli
38	35	38.9	554	4	US-09-321-276-4	Sequence 4, Appli
39	35	38.9	554	4	US-08-916-481-2	Sequence 2, Appli
40	35	38.9	563	4	US-08-916-481-3	Sequence 3, Appli
41	35	38.9	768	2	US-08-222-617A-5	Sequence 39, Appli
42	35	38.9	1097	2	US-08-680-326-39	Sequence 2, Appli
43	35	38.9	1205	1	US-07-908-245-2	Sequence 12, Appli
44	35	38.9	3666	2	US-08-222-617A-12	Sequence 27, Appli
45	35	38.9	3727	2	US-08-222-617A-27	

ALIGNMENTS

RESULT 1
US-07-906-871-14
Sequence 14, Application US/07906871
Patent No. 5340739
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1225 Connecticut Avenue, N.W., Suite 300
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk.
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/906, 871
FILING DATE: 19920103
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/816, 289
FILING DATE: 03 JAN 1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/635, 544
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/03051
FILING DATE: 13-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/224, 035
FILING DATE: 13-JUL-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cimbal, Michele A
REGISTRATION NUMBER: 33, 851
REFERENCE/DOCKET NUMBER: 0627, 2830004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)833-8716
TELEFAX: (202)833-8716
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: AMINO ACID

;; TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
;; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLYCIN AND USES
;; TITLE OF INVENTION: THEREOF
;; NUMBER OF SEQUENCES: 18
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;; STREET: 1225 Connecticut Avenue, N.W., Suite 300
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20036
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/906,871
;; FILING DATE: 19920103
;; CLASSIFICATION: 435
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/816,289
;; FILING DATE: 03 JAN 1992
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/635,544
;; FILING DATE: 18-JAN-1991
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US89/03051
;; FILING DATE: 13-JUL-1989
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/224,035
;; FILING DATE: 13-JUL-1988
;;
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cimbal, Michele A.
;; REGISTRATION NUMBER: 33,851
;; REFERENCE/DOCKET NUMBER: 0627.2830004
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)833-7533
;; TELEFAX: (202)833-8716
;;
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 16 amino acids
;; TYPE: AMINO ACID
;; STRANDEDNESS: both
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;;
;; US-07-906-871-4

Query Match 67.8%; Score 61; DB 1; Length 16;
Best Local Similarity 83.3%; Pred. No. 0.002;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YPARARYQWVR 12
DB 5 YPTQRAYQWVR 16

RESULT 5
US-08-591-632-69
Sequence 69, Application US/08591632
Patent No. 6261558
GENERAL INFORMATION:
APPLICANT: Barbados, Carlos F.
APPLICANT: Burton, Dennis R.
APPLICANT: Lerner, Richard A.
TITLE OF INVENTION: SYNTHETIC HUMAN NEUTRALIZING MONOCLONAL
TITLE OF INVENTION: ANTIBODIES TO HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESSEE: Patent Counsel
STREET: 10550 No. 6261558th Torrey Pines Road, TPC 8
CITY: La Jolla

;; STATE: CA
;; COUNTRY: USA
;; ZIP: 92037
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: Patentin Release #1.0
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/591,632
;; FILING DATE: 19-OCT-1994
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/11907
;; FILING DATE: 19-OCT-1994
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/308,841
;; FILING DATE: 19-SEP-1994
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/233,619
;; FILING DATE: 26-APR-1994
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/139,409
;; FILING DATE: 19-OCT-1993
;;
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Filling, Thomas
;; REGISTRATION NUMBER: 34,163
;; REFERENCE/DOCKET NUMBER: TSRI 332.3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 784-2937
;; TELEFAX: (619) 784-9399
;;
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 108 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;;
;; US-08-591-632-69

Query Match 44.4%; Score 40; DB 4; Length 108;
Best Local Similarity 46.7%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 YPARARYQWVR 15
DB 25 HPLHARAYQWVR 39

RESULT 6
US-09-611-451-69
Sequence 69, Application US/09611451
Patent No. 6395275
GENERAL INFORMATION:
APPLICANT: Barbados, Carlos F.
APPLICANT: Burton, Dennis R.
APPLICANT: Lerner, Richard A.
TITLE OF INVENTION: SYNTHETIC HUMAN NEUTRALIZING MONOCLONAL
TITLE OF INVENTION: ANTIBODIES TO HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESSEE: Patent Counsel
STREET: 10550 No. 6395275th Torrey Pines Road, TPC 8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/611,451

OY 1 YPARRRAR-----YQWVRCKP 15
|| : | : || : ||
Db 3469 YPEAKTRLATKVMYKWKQKP 3490

RESULT 9

US-08-750-717-2
; Sequence 2, Application US/08750717
; Patent No. 6180109
; GENERAL INFORMATION:
; APPLICANT: MOORMANN, Robertus J. M.
; APPLICANT: VAN RIJN, Petrus A.
; TITLE OF INVENTION: Nucleotide Sequences of Pestivirus
; TITLE OF INVENTION: Strains, Polypeptides Encoded by These Sequences and Use
; TITLE OF INVENTION: Therapies for Diagnosis and Prevention of Pestivirus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; City: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,717
; FILING DATE: 24-DEC-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94201743.5
; FILING DATE: 17-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/NL95/00214
; FILING DATE: 16-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: BO 39123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-750-717-2

Query Match

Best Local Similarity 43.9%; Score 39.5; DB 4; Length 3898;
Matches 8; Conservative 3; Mismatches 4; Indels 7; Gaps 1;

OY 1 YPARRRAR-----YQWVRCKP 15
|| : | : || : ||
Db 3469 YPEAKTRLATKVMYKWKQKP 3490

RESULT 10

US-09-267-177-26
; Sequence 26, Application US/09267177
; Patent No. 6287856
; GENERAL INFORMATION:
; APPLICANT: Poet, Steven E.
; APPLICANT: Ritchie, Branson W.
; APPLICANT: Niagro, Frank D.

; APPLICANT: Lukert, Phil D.
; TITLE OF INVENTION: Vaccines against Circovirus Infections
; FILE REFERENCE: 21099.0057
; CURRENT APPLICATION NUMBER: US/09/267,177
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: 60/077,890
; EARLIER FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 229
; TYPE: PRT
; ORGANISM: porcine circovirus
; US-09-267-177-26

Query Match

Best Local Similarity 43.3%; Score 39; DB 4; Length 229;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 YPARRRARYQWVR 12
|| : | : || : ||
Db 31 HPAFRNRKRWRR 42

RESULT 11

US-09-391-104-18
; Sequence 18, Application US/09391104
; Patent No. 639371
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Falduto, Michael T.
; APPLICANT: Magnuson, Scott R.
; APPLICANT: Morgan, Douglas W.
; TITLE OF INVENTION: HUMAN MATRIX METALLOPROTEASE GENE,
; TITLE OF INVENTION: PROTEINS ENCODED THEREFROM AND METHODS
; FILE REFERENCE: 6073.US.P1
; CURRENT APPLICATION NUMBER: US/09/391,104
; CURRENT FILING DATE: 1999-09-07
; PRIOR APPLICATION NUMBER: US 08/814,394
; PRIOR FILING DATE: 1997-03-11
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-391-104-18

Query Match

Best Local Similarity 42.8%; Score 38.5; DB 4; Length 508;
Matches 6; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

OY 1 YPARRRARYQWVRCKP 15
|| : | : || : ||
Db 461 YP-RNISHNMHCRP 474

RESULT 12

US-08-984-709A-50
; Sequence 50, Application US/08984709A
; Patent No. 6320032
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark E.
; APPLICANT: Stauder, Kenneth A.
; APPLICANT: Harpold, Michael M.
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McCaulliffe
; STREET: 4250 Executive Square, Suite 700
; CITY: La Jolla

STATE: California
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,709A
FILING DATE: 02-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 24735-9815 (formerly 6362-9815)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 450-8400
TELEFAX: (619) 587-5360
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 2353 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-984-709A-50

Query Match 42.2%; Score 38; DB 4; Length 2353;
Best Local Similarity 63.6%; Pred. No. 1e+03;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 4 RRARYVRCK 14
| | | | |
Db 1474 RAAHYRWVRK 1484

RESULT 13
US-08-484-105-18
Sequence 18, Application US/08484105
Patent No. 5589341
GENERAL INFORMATION:
APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: RINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: MCNALLY, Francis J
APPLICANT: LAURENSEN, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: LI, Joachim J
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,105
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 861 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-105-18

Query Match 41.1%; Score 37; DB 1; Length 861;
Best Local Similarity 66.7%; Pred. No. 5.4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 PARRARYQW 10
| | | | |
Db 80 PKRARYQW 88

RESULT 14
US-08-484-106-18
Sequence 18, Application US/08484106
Patent No. 5614618
GENERAL INFORMATION:
APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: RINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: MCNALLY, Francis J
APPLICANT: LAURENSEN, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: LI, Joachim J
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,106
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 861 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-484-106-18

Query Match

41.1%; Score 37; DB 1; Length 861;

Best Local Similarity 66.7%; Pred. No. 5.4e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 PARRARYQM 10

Db 80 PKRRARYQM 88

RESULT 15

US-09-000-145-4

; Sequence 4, Application US/09000145

; Patent No. 6169172

; GENERAL INFORMATION:

; APPLICANT: DEVAUCHELLE, Gerard

; APPLICANT: GARNIER, Laurence

; APPLICANT: CAHOREAU, Claire

; APPLICANT: CERUTTI, Martine

; TITLE OF INVENTION: USE OF A PROLACTIN RECEPTOR OR GROWTH HORMONE RECEPTOR

; TITLE OF INVENTION: INTRACYTOPLASMIC DOMAIN FOR ACHIEVING PROTEIN SECRETION

; FILE REFERENCE: 0384-0047-0XECT

; CURRENT APPLICATION NUMBER: US/09/000,145

; CURRENT FILING DATE: 1998-03-16

; EARLIER APPLICATION NUMBER: PCT/FR96/01237

; EARLIER FILING DATE: 1996-08-02

; EARLIER APPLICATION NUMBER: FR 95/09420

; EARLIER FILING DATE: 1995-08-02

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 593

; TYPE: PRT

; ORGANISM: Oryctolagus cuniculus

US-09-000-145-4

Query Match

40.6%; Score 36.5; DB 4; Length 593;

Best Local Similarity 53.3%; Pred. No. 4.4e+02;

Matches 8; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 1 YPARARYQWVRCKP 15

Db 174 YPGOKYLVQ-VRCKP 187

Search completed: April 8, 2003, 11:31:16
Job time : 21.6552 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:11:45 ; Search time 33.7931 seconds
(Without alignments)
31.545 Million cell updates/sec

Title: US-09-496-391-15
Perfect score: 36
Sequence: 1 AKAARRRA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	8	21	AAB08138
2	33	91.7	24	21	AAB08162
3	31	86.1	215	22	AAU53203
4	31	86.1	734	21	AA632174
5	31	86.1	751	21	AA639904
6	31	86.1	764	21	AA632173
7	31	86.1	781	21	AA639903
8	31	86.1	820	21	AA632172
9	31	86.1	837	21	AA639902
10	30	83.3	14	21	AA621902

11	30	83.3	14	22	AAE06480	Synthetic anti-mic
12	30	83.3	14	23	AAU98815	Pro-apoptosis pept
13	30	83.3	14	23	ABG60328	Selective targetin
14	30	83.3	14	23	ABG34927	Pro-apoptosis pept
15	30	83.3	18	17	AAW07280	Amphiphilic antim
16	30	83.3	18	22	AAW74628	18 mer peptide seq
17	30	83.3	21	22	AAW74630	21 mer peptide seq
18	30	83.3	24	21	AAW08171	Peptide modulating
19	30	83.3	24	21	AAW08178	Peptide modulating
20	30	83.3	32	21	AAB08168	Peptide modulating
21	30	83.3	40	21	AAB08170	Peptide modulating
22	30	83.3	101	22	AAU33670	Pseudomonas aerugi
23	30	83.3	145	18	AAW22489	Phaffia derived gl
24	30	83.3	198	22	AAW22489	Human protein sequ
25	30	83.3	227	20	AAW38859	Neisseria meningit
26	30	83.3	227	23	AAU73006	Human DNA modifia
27	30	83.3	279	22	AAU03592	Human protein sequ
28	30	83.3	454	21	AAW06059	Arabidopsis thalia
29	30	83.3	456	22	AAW22489	Arabidopsis thalia
30	30	83.3	467	22	AAW22489	Human protein sequ
31	30	83.3	548	21	AAW06058	Arabidopsis thalia
32	30	83.3	550	21	AAW06058	Arabidopsis thalia
33	30	83.3	558	21	AAW06057	Arabidopsis thalia
34	30	83.3	560	21	AAW06057	Arabidopsis thalia
35	30	83.3	650	22	AAW06057	Arabidopsis thalia
36	29	80.6	14	21	AAW06057	Arabidopsis thalia
37	29	80.6	52	23	AAW06057	Human ORF protein
38	29	80.6	87	22	AAW06057	Escherichia coli p
39	29	80.6	130	23	AAU09945	Peptide fragment o
40	29	80.6	130	23	AAE13237	Human linker histo
41	29	80.6	238	21	AAW06057	Arabidopsis thalia
42	29	80.6	238	21	AAW06057	Arabidopsis thalia
43	29	80.6	339	21	AAW06057	Arabidopsis thalia
44	29	80.6	367	22	AAW06057	Arabidopsis thalia
45	29	80.6	427	22	AAW06057	Protonibacterium

ALIGNMENTS

RESULT 1	
AAB08138	
ID	AAW08138 standard; peptide: 8 AA.
XX	
AC	AAW08138;
XX	
DT	04-DEC-2000 (first entry)
DE	
DE	Peptide modulating activity of heparin, and other glycans.
XX	
KW	Glycosaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW	cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW	cartilage differentiation; wound healing.
XX	
OS	Synthetic.
XX	
FT	Key
FT	Misc-difference 1.8
FT	/note="this peptide may be repeated an
FT	unspecified number of times"
XX	
PN	WO200045831-A1.
XX	
PD	10-AUG-2000.
XX	
PF	02-FEB-2000; 2000WO-USO2853.
XX	
PR	02-FEB-1999; 99US-0118276.
XX	
PA	(UYJF-) UNIV JEFFERSON THOMAS.
XX	
PI	San Antonio JD, Verrecchio A, Schick BP;

DR WPI; 2000-543446/49.
XX
XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
XX
PS Disclosure; Page 23; 76pp; English.
XX
XX The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with
CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
CC
XX
SQ Sequence 8 AA;
Query Match 100.0%; Score 36; DB 21; Length 8;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKAARRA 8
DB 1 AKAARRA 8
IIIIIIII
RESULT 2
AAB08162
ID AAB08162 standard; peptide: 24 AA.
XX
XX AAB08162;
AC
XX
DT 04-DEC-2000 (first entry)
XX
XX Peptide modulating activity of heparin, and other glycans.
DE
XX Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW cartilage differentiation; wound healing.
XX
OS Synthetic.
XX
XX WO200045831-A1.
PN
XX
XX 10-AUG-2000.
PD
XX
XX 02-FEB-2000; 2000WO-US02853.
PF
XX
XX 02-FEB-1999; 99US-0118276.
PR
XX
XX (UYJE-) UNIV JEFFERSON THOMAS.
PA
XX
PI San Antonio JD, Verrecchio A, Schlick BP;
XX
XX WPI; 2000-543446/49.
DR
XX
XX Novel synthetic peptides with high affinity for glycoaminoglycans and
PT proteoglycans, useful for modulating heparin, promoting cell
PT attachment, modulating tumour metastasis and modulating wound healing -
XX
XX
PS Disclosure; Page 24; 76pp; English.
XX
XX The present sequence represents a synthetic peptide which has a high
CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
CC in methods for modulating heparin or other glycoaminoglycans with

CC anticoagulant activity, promoting cell attachment or adhesion to
CC natural or synthetic surfaces (especially vein grafts), modulating
CC tumour cell metastasis, modulating cartilage differentiation, targeting
CC drugs to epithelial cell surfaces (or to other cells expressing
CC proteoglycans), modulating enzymes that act on glycoaminoglycan
CC substrates, affinity purification of bioactive sequences of a
CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
CC anti-coagulant functions mediated through glycoaminoglycans, and
CC modulating wound healing. The peptide may also be used for blocking
CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
CC increase heparin half-life in circulation.
CC
XX
SQ Sequence 24 AA;
Query Match 91.7%; Score 33; DB 21; Length 24;
Best Local Similarity 87.5%; Pred. No. 8.1;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AKAARRA 8
DB 9 AKAARRA 16
IIIIIIII
RESULT 3
AAU53203
ID AAU53203 standard; Protein; 215 AA.
XX
XX AAU53203;
AC
XX
DT 27-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #14099.
DE
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
XX
OS Propionibacterium acnes.
XX
XX
XX WO200181581-A2.
PN
XX
XX 01-NOV-2001.
PD
XX
XX 20-APR-2001; 2001WO-US12865.
PF
XX
XX 21-APR-2000; 2000US-199047P.
PR
XX
XX 02-JUN-2000; 2000US-208841P.
PR
XX
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
DR
XX
XX N-PSDB; AAS59559.
PT
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX
XX Example 1; SEQ ID No 14398; 1069pp; English.
PS
XX
XX Sequences AAU9105-AAU6017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a

CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 215 AA;

Query Match 86.1%; Score 31; DB 22; Length 215;
Best Local Similarity 87.5%; Pred No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
DB 161 AKAKKNA 168

RESULT 4
AAG32174
ID AAG32174 standard; Protein: 734 AA.
XX
AC AAG32174;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 38763.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KM hydridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132408.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 86.1%; Score 31; DB 21; Length 734;
Best Local Similarity 87.5%; Pred. No. 4..4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AKAKKRA 8
Db 619 AKAKKRA 626

RESULT 5
ID AAG39904 standard; Protein; 751 AA.
AC AAG39904;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49440.
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142360.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159332.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 86.1%; Score 31; DB 21; Length 751;
Best Local Similarity 87.5%; Pred. No. 4.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
||| ||||
Db 635 AKAKKRA 642

RESULT 6
AAG32173
ID AAG32173 standard; Protein: 764 AA.
XX AAG32173;
AC
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 38762.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0137822.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153738.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 86.1%; Score 31; DB 21; Length 764;
Best Local Similarity 87.5%; Pred. No. 4.6e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ARAAKRA 8
DB 649 ARAAKRA 656

RESULT 7
AAG39903
ID AAG39903 standard; Protein: 781 AA.
AC AAG39903;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 49439.
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
OS Arabidopsis thaliana.
PN EPI33405-A2.
XX 06-SEP-2000.
PD 25-FEB-2000; 200EP-0301439.
PF 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 30-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144884.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145813.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147935.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.

PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151066.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151308.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 86.1% Score 31: DB 21: Length 781:
Best Local Similarity 87.5% Pred. No. 4.7e+02:
Matches 7: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 1 AKAKKRA 8
Db 665 AKAKKRA 672

RESULT 8
AAG32172
ID AAG32172 standard: Protein: 820 AA.
XX
AC AAG32172:

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 86.1%; Score 31; DB 21; Length 820;

Best Local Similarity 87.5%; Pred. No. 4.9e+02; Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKAKKRA 8
Db 705 AKAKKRA 712

RESULT 9
AAG39902
ID AAG39902 standard; Protein; 837 AA.

XX AAG39902;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 49438.

KW Protein identification: signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.
XX Arabidopsis thaliana.
OS EP1033405-A2.
PN 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
PF 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 86.1%; Score 31; DB 21; Length 837;
Best Local Similarity 87.5%; Pred. No. 5e+02; 1; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 1;

QY 1 AKAKKRA 8
DB 721 AKAKKRA 728

RESULT 10
AAB21902
ID AAB21902 standard; peptide; 14 AA.

AC AAB21902;

DT 22-MAR-2001 (first entry)

DE Antimicrobial pro-apoptotic peptide #3.

KW Cytostatic; homing pro-apoptotic conjugate; tumour; antimicrobial;
KW breast; prostate; melanoma; cancer; Kaposi's sarcoma; amphipathic;
KW alpha-helix.

OS Unidentified.

PN WO200042973-A2.

PD 27-JUL-2000.

XX 21-JAN-2000; 2000WO-US01602.
XX
XX 22-JAN-1999; 99US-0235902.
XX
XX (BURN-) BURNHAM INST.
XX
XX Ellerby HM, Bredesen DE, Pasqualini R, Ruoslahti EI;
XX WPI; 2000-499174/44.
XX
XX Homing pro-apoptotic conjugate comprising a tumor homing molecule that
PT selectively homes to a mammalian cell type or tissue linked to an
PT antimicrobial peptide; useful for the treatment of prostate cancer -
XX
XX Claim 4; Page 104; 118pp; English.
XX
XX The present invention relates to homing pro-apoptotic conjugates,
CC comprising of a tumor homing molecule that selectively homes to a
CC mammalian cell type or tissue, linked to an antimicrobial peptide. The
CC homing pro-apoptotic conjugates are selectively internalised by the
CC mammalian cell type or tissue and exhibits high toxicity, especially to
CC angiogenic vasculature. The antimicrobial peptide has low mammalian cell
CC toxicity when not linked to the tumor homing molecule. The conjugates are
CC useful for the treatment of cancer e.g. Kaposi's sarcoma, breast and
CC prostate cancer or melanoma. The present sequence is one such
CC antimicrobial peptide, which can be conjugated to a homing peptide to
CC make the homing pro-apoptotic conjugates of the present invention. The
CC present sequence has an amphipathic alpha-helical structure.
XX
SQ Sequence 14 AA;
XX
Query Match 83.3%; Score 30; DB 21; Length 14;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 AKAKKRA 8
Db 7 AKAKKRA 14
XX
RESULT 11
AAE06480
ID AAE06480 standard; peptide; 14 AA.
XX
XX AAE06480;
XX
XX 25-SEP-2001 (first entry)
DT
XX
XX Synthetic anti-microbial peptide #3.
DE
XX
XX Chimeric prostate-homing pro-apoptotic peptide; prostate-homing peptide;
KW antimicrobial peptide; prostate cancer; breast tumor homing molecule;
KW cytostatic.
XX
XX Synthetic.
OS
XX
XX WO200153342-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 16-JAN-2001; 2001WO-US01362.
PF
XX
XX 21-JAN-2000; 2000US-0489582.
PR
XX
XX (BURN-) BURNHAM INST.
PA
XX
XX Ruoslahti EI, Pasqualini R, Arap W, Bredesen DE, Ellerby HM;
PI WPI; 2001-451901/48.
DR
XX
XX Novel chimeric prostate-homing pro-apoptotic peptide; used to treat
PT prostate cancer, comprises a prostate-homing peptide linked to an

PT antimicrobial peptide -
XX
XX Claim 4; Page 103; 176pp; English.
XX
XX The patent discloses novel chimeric prostate-homing pro-apoptotic
CC peptide which comprises a prostate-homing peptide linked to an
CC antimicrobial peptide, where the chimeric peptide is selectively
CC internalised by and exhibits high toxicity to prostate tissue and
CC where the antimicrobial peptide has low mammalian cell toxicity when
CC not linked to prostate-homing peptide. The chimeric peptide is used
CC to direct an antimicrobial peptide in vivo to a prostate cancer, to
CC induce selective toxicity in vivo in a prostate cancer, and to treat
CC a patient with prostate cancer. The present sequence is a synthetic
CC anti-microbial peptide having an amphipathic-alpha helical structure.
CC This peptide is linked to a prostate-homing peptide to generate a
CC chimeric prostate-homing pro-apoptotic peptide.
XX
SQ Sequence 14 AA;
XX
Query Match 83.3%; Score 30; DB 22; Length 14;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 AKAKKRA 8
Db 7 AKAKKRA 14
XX
RESULT 12
AAU98815
ID AAU98815 standard; Peptide; 14 AA.
XX
XX AAU98815;
AC
XX
XX 22-AUG-2002 (first entry)
DT
XX
XX Pro-apoptosis peptide #3.
DE
XX
XX Targeting peptide; cancer; cytostatic; anti-HIV,
KW immunostimulant; immunogen; cancer; human immunodeficiency virus;
KW HIV; vector delivery; apoptosis.
XX
XX Synthetic.
OS
XX
XX WO200220724-A2.
PN
XX
XX 14-MAR-2002.
PD
XX
XX 07-SEP-2001; 2001WO-US28045.
PF
XX
XX 08-SEP-2000; 2000US-231266P.
PR
XX
XX 17-JAN-2001; 2001US-0765101.
PR
XX
XX (TEXA) UNIV TEXAS SYSTEM.
PA
XX
XX Arap W, Pasqualini R;
PI WPI; 2002-489672/52.
DR
XX
XX Modulation of immune system response comprises administration of
PT targeting peptide attached to immunogen -
XX
XX Disclosure; Page 14; 86pp; English.
XX
XX This invention relates to a method for modulating the immune system
CC response comprising administration of a lymph node targeting peptide
CC attached to an immunogen. The invention also comprises a bispecific
CC compound comprising the sequences Cys-Ala-Tyr or Ser-Cys-Ala-Arg,
CC a bispecific compound comprising a targeting peptide attached to a
CC vector binding moiety and a method for targeting a vector to an organ or
CC tissue comprising administering the vector and a complex comprising a
CC targeting peptide and a binding moiety. The peptides of the invention
CC may have cytostatic, anti-HIV or immunostimulant activities. The method

CC of the invention is useful for increasing the immune response to an
CC immunogen, especially a cancer cell or human immunodeficiency virus
CC (HIV). The method is useful for the selective delivery of gene
CC therapy vectors. The present sequence represents a pro-apoptotic
CC peptide used in the method of the invention.

XX Sequence 14 AA:

Query Match 83.3%; Score 30; DB 23; Length 14;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARRA 8
XX |||||
Db 7 AKAARRA 14

RESULT 13
ABG60328
ID ABG60328 standard; Peptide: 14 AA.

AC ABG60328;

XX 30-JUL-2002 (first entry)

DE Selective targeting peptide #3.

XX Targeting peptide: cancer; arthritis; diabetes; inflammatory disease;
KW atherosclerosis; autoimmune disease; bacterial infection; apoptosis;
KW viral infection; cardiovascular disease; degenerative disease; ischaemia;
KW inflammation; macular degeneration; anti-inflammatory; antidiabetic;
KW cardiovascular; immunomodulator; antibacterial; antiviral; cytostatic;
KW gene therapy.

XX Synthetic.

PN WO20020769-A1.

PD 14-MAR-2002.

PE 07-SEP-2001; 2001WO-US27692.

PR 08-SEP-2000; 2000US-231266P.

PR 17-JAN-2001; 2001US-0765101.

XX (TEXA) UNIV TEXAS SYSTEM.

PI Arap W, Pasqualini R;

DR WPI; 2002-415731/44.

PT Targeting peptides identified by phage display, useful for targeting
PT delivery to an organ or tissue, particularly for treating a disease,
PT e.g. cancer, inflammatory or autoimmune diseases, infections or
PT cardiovascular disease -

PS Claim 31; Page 149; 317pp; English.

XX The invention relates to an isolated peptide of 100 amino acids or less
CC in size useful for targeting delivery to an organ or tissue, particularly
CC for treating a disease, e.g. cancer, arthritis, diabetes, inflammatory
CC disease, atherosclerosis, autoimmune disease, bacterial infection, viral
CC infection, cardiovascular disease or degenerative disease. The peptide is
CC also useful for inducing apoptosis, particularly to a subject with
CC ischaemia, cancer, arthritis, diabetes, cardiovascular disease,
CC inflammation or macular degeneration. Furthermore, the peptide is useful
CC for diagnosing the diseases cited above. Targeting peptides of the
CC invention can also be used to deliver an agent to a foetus, by attaching
CC a peptide to the agent and administering the peptide to a pregnant
CC subject. Sequences ABG60326-ABG60574 represent selective targeting
CC peptides of the invention.

XX Sequence 14 AA:

Query Match 83.3%; Score 30; DB 23; Length 14;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARRA 8
XX |||||
Db 7 AKAARRA 14

RESULT 14
ABG34927
ID ABG34927 standard; Peptide: 14 AA.

AC ABG34927;

DE 15-JUL-2002 (first entry)

DE Pro-apoptosis peptide agent #3.

XX Targeting peptide: cancer; Hodgkin's disease; cytostatic;
KW immunosuppressive; anti-inflammatory; antiarthritic; antiviral;
KW antiatherosclerotic; antidiabetic; antibacterial; diabetes mellitus;
KW inflammatory disease; arthritis; atherosclerosis; cancer;
KW autoimmune disease; bacterial infection; viral infection.

XX Unidentified.

PN WO20020722-A2.

PD 14-MAR-2002.

PE 07-SEP-2001; 2001WO-US27702.

PR 08-SEP-2000; 2000US-231266P.

PR 17-JAN-2001; 2001US-0765101.

XX (TEXA) UNIV TEXAS SYSTEM.

PI Arap W, Pasqualini R;

DR WPI; 2002-383050/41.

PT Identifying targeting peptides useful for treating e.g. diabetes
PT mellitus, inflammatory diseases, cancer, or autoimmune diseases,
PT comprises exposing a sample to a phage display library and recovering
PT phage bound to the sample -

PS Claim 67; Page 200; 298pp; English.

XX This invention relates to a novel method for identifying disease
CC targeting peptides. The method comprises exposing a sample from an
CC organ, tissue or cell type of interest, to a phage display library and
CC recovering phage bound to the sample (the phage expresses targeting
CC peptides). The peptides identified by the method of the invention may
CC have cytostatic, immunosuppressive, anti-inflammatory, antiarthritic,
CC antiatherosclerotic, antidiabetic, antibacterial and antiviral
CC activities. The methods and composition are useful for identifying
CC targeting peptides and one or more receptors for a targeting peptide.
CC The targeting peptides are used for selective delivery of therapeutic
CC agents, including gene therapy vectors and fusion proteins, to specific
CC organs, tissues, or cell types in subject. The targeting peptide may
CC also be used for treating diseases such as diabetes mellitus,
CC inflammatory diseases, arthritis, atherosclerosis, cancer, autoimmune
CC diseases, bacterial and viral infections and Hodgkin's disease. The
CC present sequence represents a targeting peptide of the invention.

XX Sequence 14 AA:

Query Match 83.3%; Score 30; DB 23; Length 14;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARRA 8

Job time : 34.7931 secs

DB 7 AKAARRA 14

RESULT 15

ID AAW07290 standard; peptide; 18 AA.

AC AAW07290;

DT 29-MAY-1997 (first entry)

DE Amphiphilic antimicrobial peptide MB-50.

KW amphiphilic antimicrobial; log kill; Staphylococcus aureus; charge;

KW Escherichia coli; vectorial analysis; hydrophile; lipophile balance;

KW hydrophobic moment; equation; antimicrobial; antibacterial;

XX antifungal; disinfection; spoilage prevention; preservation.

OS Synthetic.

XX WO9628468-A2.

XX 19-SEP-1996.

XX 27-FEB-1996; 96WO-EP00844.

XX 09-MAR-1995; 95GB-0004761.

XX (UNIL) UNILEVER NV.

XX (UNIL) UNILEVER PLC.

XX Bhakoo M, Patel S, Stott PI;

XX WPI; 1996-433760/43.

XX New amphiphilic antimicrobial peptide(s) - having particular

XX combinations of hydrophilic and hydrophobic amino acid residues

XX Example -: Page 22; 39pp; English.

XX AAW07270-90 examples of amphiphilic antimicrobial peptides for

XX comparison to claimed amphiphilic antimicrobial peptides (AAW07257-69),

XX and show that relatively minor changes in structure can have significant

XX consequences as regards the antimicrobial activity of the molecules. The

XX log kills were predicted against S. aureus and against E. coli, and

XX values of more than 5 and 4 respectively, correspond to greater

XX antimicrobial activity. Effective antimicrobial peptides are

XX discriminated from ineffective peptides by a vectorial analysis on

XX dimensions corresponding to charge, hydrophile/lipophile balance,

XX hydrophobic moment and amphiphilicity of the peptides, where effective

XX peptides fall into the region which is anti-clockwise of the

XX amphiphilicity dimension and clockwise of the charge dimension.

XX Effective peptides are further discriminated from ineffective peptides by

XX means of an equation relating certain physical properties of the peptides

XX to their biological activity against specific microorganisms. The

XX peptides can be used in antimicrobial, pref. antibacterial or antifungal,

XX compsns. They can be used for the disinfection of surfaces, spoilage

XX prevention, preservation or other hygiene processes.

SQ Sequence 18 AA;

Query Match 83.3%; Score 30; DB 17; Length 18;

Best Local Similarity 87.5%; Pred. No. 24;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARRA 8

DB 5 AKAARRA 12

Search completed: April 8, 2003, 11:23:47

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:19:38 ; Search time 10.4828 seconds
(without alignments)
22.454 Million cell updates/sec

Title: US-09-496-391-15
Perfect score: 36
Sequence: 1 AKAKKRA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_Aa: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PTOUTS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	83.3	145	4	US-09-091-725-37 Sequence 37, Appl
2	29	80.6	45	4	US-08-981-392-10 Sequence 10, Appl
3	29	80.6	433	1	US-08-346-849-2 Sequence 2, Appl
4	29	80.6	433	2	US-08-293-284A-2 Sequence 2, Appl
5	28	77.8	280	4	US-09-145-828A-21 Sequence 21, Appl
6	28	77.8	293	4	US-09-145-828A-17 Sequence 17, Appl
7	28	77.8	317	4	US-09-145-828A-7 Sequence 12, Appl
8	28	77.8	317	4	US-09-145-828A-7 Sequence 12, Appl
9	28	77.8	317	4	US-09-145-828A-7 Sequence 12, Appl
10	28	77.8	1185	4	US-09-041-886-23 Sequence 19, Appl
11	27	75.0	24	2	US-08-491-527A-13 Sequence 13, Appl
12	27	75.0	208	4	US-09-134-001C-4971 Sequence 4971, Ap
13	27	75.0	325	3	US-08-581-148C-4 Sequence 4, Appl
14	27	75.0	346	4	US-09-352-990-23 Sequence 23, Appl
15	27	75.0	655	4	US-08-857-076-57 Sequence 57, Appl
16	27	75.0	885	4	US-09-342-648-9 Sequence 9, Appl
17	26	72.2	19	2	US-08-660-592-5 Sequence 5, Appl
18	26	72.2	19	4	US-09-166-930A-5 Sequence 5, Appl
19	26	72.2	109	4	US-09-341-444A-2 Sequence 5, Appl
20	26	72.2	138	4	US-09-134-001C-4254 Sequence 4254, Ap
21	26	72.2	214	3	US-09-041-889-27 Sequence 27, Appl
22	26	72.2	368	2	US-08-651-579-2 Sequence 2, Appl
23	26	72.2	395	4	US-08-900-230-5 Sequence 5, Appl
24	26	72.2	492	3	US-09-006-636-4 Sequence 4, Appl
25	26	72.2	492	4	US-09-006-632-4 Sequence 4, Appl
26	26	72.2	492	4	US-09-325-274-4 Sequence 4, Appl
27	25	69.4	22	3	US-08-940-095-105 Sequence 105, App

28	25	69.4	22	3	US-08-940-093-105	Sequence 105, App
29	25	69.4	22	3	US-08-940-096-105	Sequence 105, App
30	25	69.4	22	4	US-09-465-719-105	Sequence 105, App
31	25	69.4	22	4	US-09-453-605-105	Sequence 105, App
32	25	69.4	22	4	US-09-453-838-105	Sequence 105, App
33	25	69.4	27	4	US-07-946-180B-4	Sequence 4, Appl
34	25	69.4	28	1	US-08-303-025-12	Sequence 12, Appl
35	25	69.4	28	2	US-08-436-703B-1	Sequence 1, Appl
36	25	69.4	29	1	US-08-152-488-10	Sequence 10, Appl
37	25	69.4	29	1	US-08-152-488-11	Sequence 11, Appl
38	25	69.4	29	1	US-08-152-488-12	Sequence 12, Appl
39	25	69.4	29	1	US-08-303-025-10	Sequence 10, Appl
40	25	69.4	29	1	US-08-303-025-11	Sequence 11, Appl
41	25	69.4	29	1	US-08-303-025-13	Sequence 13, Appl
42	25	69.4	29	1	US-08-303-025-14	Sequence 14, Appl
43	25	69.4	29	1	US-08-677-304-10	Sequence 10, Appl
44	25	69.4	29	1	US-08-677-304-11	Sequence 11, Appl
45	25	69.4	29	1	US-08-677-304-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-091-725-37
; Sequence 37, Application US/09091725
; Patent No. 6329141
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Improved methods for transforming Plaffia
; TITLE OF INVENTION: and recombinant DNA for use therein
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster LLP
; STREET: 2000 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: United States of America
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,725
; FILING DATE: 23-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95203620.0
; FILING DATE: 22-DEC-1995
; APPLICATION NUMBER: EP 96200943.7
; FILING DATE: 11-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: E. Victor Donahue
; REGISTRATION NUMBER: 35,492
; INFORMATION FOR SEQ. ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-091-725-37

Query Match 83.3%; Score 30, DB 4; Length 145;
Best Local Similarity 87.5%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKAKKRA 8
DB 20 AKAKKAA 27

RESULT 2
US-08-981-392-10
Sequence 10, Application US/08981392
Patent No. 6262025
GENERAL INFORMATION:
APPLICANT: Ish-Horowitz, David
APPLICANT: Henrique, Domingos Manuel Pinto
APPLICANT: Lewis, Julian Hart
APPLICANT: Artavanis-Tsakonas, Spyridon
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES
TITLE OF INVENTION: OF VERTEBRATE DELTA GENES AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/981,392
FILING DATE: 22-DEC-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Antler, Adriane M.
REGISTRATION NUMBER: 32,605
REFERENCE/DOCKET NUMBER: 7326-038
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-981-392-10
Query Match 80.6%; Score 29; DB 4; Length 45;
Best Local Similarity 85.7%; Pred. No. 52;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKAARR 7
|||||
DB 21 AKAARR 27
RESULT 3
US-08-346-849-2
Sequence 2, Application US/08346849
Patent No. 5670483
GENERAL INFORMATION:
APPLICANT: Zhang, Shuangang
APPLICANT: Lockshin, Curtis
APPLICANT: Rich, Alexander
APPLICANT: Holmes, Todd
TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts

COUNTRY: U.S.A.
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/346,849
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,326
FILING DATE: 28 DECEMBER 1992
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: MIT-6008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-346-849-2
Query Match 80.6%; Score 29; DB 1; Length 433;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AKAARR 7
|||||
DB 347 AKAARR 353
RESULT 4
US-08-293-284A-2
Sequence 2, Application US/08293284A
Patent No. 5955343
GENERAL INFORMATION:
APPLICANT: Holmes, Todd
APPLICANT: Zhang, Shuangang
APPLICANT: Rich, Alexander
APPLICANT: DiPersio, C. Michael
APPLICANT: Lockshin, Curtis
TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/293,284A
FILING DATE: 22-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,326
FILING DATE: 28-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.

REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: MIT-6008A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-293-284A-2

Query Match 80.6%; Score 29; DB 2; Length 433;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKR 7
|||||:
Db 347 AKAARKK 353

RESULT 5
US-09-145-828A-21
Sequence 21, Application US/09145828A
Patent No. 6403349
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Mukerji, Pradip
APPLICANT: Leonard, Amanda E. Y.
APPLICANT: Huang, Yung-Sheng
APPLICANT: Thurmond, Jennifer
APPLICANT: Kirchner, Stephen J.
TITLE OF INVENTION: THE ELONGASE GENE AND USES THEREOF
FILE REFERENCE: 6407.US.O1
CURRENT APPLICATION NUMBER: US/09/145,828A
CURRENT FILING DATE: 1998-09-02
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 280
TYPE: PRT
ORGANISM: Mortierella alpina
FEATURE:
NAME/KEY: VARIANT
LOCATION: (280)...(280)
OTHER INFORMATION: Xaa = Unknown or other at position 280
US-09-145-828A-21

Query Match 77.8%; Score 28; DB 4; Length 280;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKR 7
|||||:
Db 249 AKAARK 255

RESULT 6
US-09-145-828A-17
Sequence 17, Application US/09145828A
Patent No. 6403349
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Mukerji, Pradip
APPLICANT: Leonard, Amanda E. Y.
APPLICANT: Huang, Yung-Sheng
APPLICANT: Thurmond, Jennifer
APPLICANT: Kirchner, Stephen J.
APPLICANT: Parker-Barnes, Jennifer M.
TITLE OF INVENTION: THE ELONGASE GENE AND USES THEREOF
FILE REFERENCE: 6407.US.O1

CURRENT APPLICATION NUMBER: US/09/145,828A
CURRENT FILING DATE: 1998-09-02
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: PRT
ORGANISM: Mortierella alpina
FEATURE:
NAME/KEY: VARIANT
LOCATION: (289)...(289)
OTHER INFORMATION: Xaa = Unknown or other at position 289
US-09-145-828A-17

Query Match 77.8%; Score 28; DB 4; Length 289;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKR 7
|||||:
Db 258 AKAARK 264

RESULT 7
US-09-145-828A-12
Sequence 12, Application US/09145828A
Patent No. 6403349
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Mukerji, Pradip
APPLICANT: Leonard, Amanda E. Y.
APPLICANT: Huang, Yung-Sheng
APPLICANT: Thurmond, Jennifer
APPLICANT: Kirchner, Stephen J.
APPLICANT: Parker-Barnes, Jennifer M.
TITLE OF INVENTION: THE ELONGASE GENE AND USES THEREOF
FILE REFERENCE: 6407.US.O1
CURRENT APPLICATION NUMBER: US/09/145,828A
CURRENT FILING DATE: 1998-09-02
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 293
TYPE: PRT
ORGANISM: Mortierella alpina
FEATURE:
NAME/KEY: VARIANT
LOCATION: (293)...(293)
OTHER INFORMATION: Xaa = Unknown or other at position 293
US-09-145-828A-12

Query Match 77.8%; Score 28; DB 4; Length 293;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKR 7
|||||:
Db 262 AKAARK 268

RESULT 8
US-09-145-828A-7
Sequence 7, Application US/09145828A
Patent No. 6403349
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Mukerji, Pradip
APPLICANT: Leonard, Amanda E. Y.
APPLICANT: Huang, Yung-Sheng
APPLICANT: Thurmond, Jennifer
APPLICANT: Kirchner, Stephen J.
APPLICANT: Parker-Barnes, Jennifer M.
TITLE OF INVENTION: THE ELONGASE GENE AND USES THEREOF

```

; FILE REFERENCE: 6407.US.01
; CURRENT APPLICATION NUMBER: US/09/145,828A
; CURRENT FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 317
; TYPE: PRF
; ORGANISM: Mortierella alpina
; US-09-145-828A-7

Query Match
Best Local Similarity 77.8%; Score 28; DB 4; Length 317;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKAARR 7
Db 287 AKAKKER 293

RESULT 9
US-09-145-828A-19
; Sequence 19, Application US/09145828A
; Patent No. 6403349
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Mukerji, Pradip
; APPLICANT: Leonard, Amanda E. Y.
; APPLICANT: Huang, Yung-Sheng
; APPLICANT: Thurmond, Jennifer
; APPLICANT: Kirchner, Stephen J.
; APPLICANT: Parker-Barnes, Jennifer M.
; TITLE OF INVENTION: THE ELONGASE GENE AND USES THEREOF
; FILE REFERENCE: 6407.US.01
; CURRENT APPLICATION NUMBER: US/09/145,828A
; CURRENT FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 318
; TYPE: PRF
; ORGANISM: Mortierella alpina
; NAME/KEY: VARIANT
; LOCATION: (318)...(318)
; OTHER INFORMATION: Xaa - Unknown or other at position 318
; US-09-145-828A-19

Query Match
Best Local Similarity 77.8%; Score 28; DB 4; Length 318;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKAARR 7
Db 287 AKAKKER 293

RESULT 10
US-09-041-886-23
; Sequence 23, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Razladeh, Sharrooz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; NUMBER OF SEQUENCES: Polypeptides and Methods of use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
```

```

; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-041-886-23

Query Match
Best Local Similarity 77.8%; Score 28; DB 4; Length 1185;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKAARR 8
Db 785 SKLAKRA 792

RESULT 11
US-08-491-527A-13
; Sequence 13, Application US/08491527A
; Patent No. 3824483
; GENERAL INFORMATION:
; APPLICANT: Houston, Michael E.
; APPLICANT: Hodges, Robert S.
; TITLE OF INVENTION: Conformationally-Restricted Combinatorial
; NUMBER OF SEQUENCES: Library Composition and Method
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,527A
; FILING DATE: 16-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/260,199
; FILING DATE: 15-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 7900-0008.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 13:
```



```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 24 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
: MOLECULE TYPE: peptide
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: single-stranded peptide, Fig. 9
: US-08-491-527A-13

```

```

Query Match
Best Local Similarity 75.0%; Score 27; DB 2; Length 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 AKAARKA 8
Db 2 AEAARKA 9

```

```

RESULT 12
US-09-134-001C-4971
: Sequence 4971, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: FILE REFERENCE: GFC-007
: CURRENT APPLICATION NUMBER: US/09/134, 001C
: PRIOR FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 4971
: LENGTH: 208
: TYPE: PRT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4971

```

```

Query Match
Best Local Similarity 75.0%; Score 27; DB 4; Length 208;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AKAARK 6
Db 46 AKAARK 51

```

```

RESULT 13
US-08-581-148C-4
: Sequence 4, Application US/08581148C
: Patent No. 6060644
: GENERAL INFORMATION:
: APPLICANT: Schnable, Patrick S.
: APPLICANT: Robertson, Donald S.
: APPLICANT: Hansen, Joel D.
: APPLICANT: Nikolau, Basil J.
: APPLICANT: Xu, Xiaojie
: APPLICANT: Xia, Yiji
: TITLE OF INVENTION: ISOLATION AND USE OF CUTICULAR LIPID
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Leydig, Voit & Mayer, Ltd.
: STREET: Two Prudential Plaza, Suite 4900
: CITY: Chicago
: STATE: IL
: COUNTRY: US
: ZIP: 60601-6780

```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/581,148C
: FILING DATE: 29-DEC-1995
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Latcher, Carol
: REGISTRATION NUMBER: 35243
: REFERENCE/DOCKET NUMBER: 71380
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 616-5600
: TELEFAX: (312) 616-5700
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 325 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
US-08-581-148C-4

```

```

Query Match
Best Local Similarity 75.0%; Score 27; DB 3; Length 325;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 AKAARKA 8
Db 317 AKDAKKA 324

```

```

RESULT 14
US-09-352-990-23
: Sequence 23, Application US/09352990
: Patent No. 6255090
: GENERAL INFORMATION:
: APPLICANT: Famodu, Layo O.
: APPLICANT: Orozco, Buddy
: APPLICANT: Rafalski, Antoni
: TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
: FILE REFERENCE: BB-1191
: CURRENT APPLICATION NUMBER: US/09/352,990
: CURRENT FILING DATE: 1999-07-14
: EARLIER APPLICATION NUMBER: 60/092,866
: EARLIER FILING DATE: July 15, 1998
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 23
: LENGTH: 346
: TYPE: PRT
: ORGANISM: Drosophila melanogaster
US-09-352-990-23

```

```

Query Match
Best Local Similarity 75.0%; Score 27; DB 4; Length 346;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 AKAARK 6
Db 17 AKAARK 22

```

```

RESULT 15
US-08-857-076-57
: Sequence 57, Application US/08857076C
: Patent No. 6225120
: GENERAL INFORMATION:
: APPLICANT: Ruwkun, Gary
: APPLICANT: Kimura, Koutarou
: APPLICANT: Patterson, Garth
: APPLICANT: Ogy, Scott

```

```

; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-857-076-57

```

```

Query Match          75.0%; Score 27; DB 4; Length 655;
Best Local Similarity 62.5%; Pred. No. 1.1e+03;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AKAKKRA 8
       :||||:|
Db      268 SRAKKKA 275

```

```

Search completed: April 8, 2003, 11:31:19
Job time : 11.4828 secs

```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:28:39 ; Search time 11.0345 Seconds
(without alignments)
44.324 Million cell updates/sec

Title: US-09-496-391-15

Perfect score: 36

Sequence: 1 AKAAKRAA 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

248812

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCIT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCITUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	83.3	14	US-09-765-086-202	Sequence 202, App
2	30	83.3	101	US-09-815-242-5166	Sequence 5166, Ap
3	29	80.6	45	US-09-908-322-10	Sequence 10, Appl
4	29	80.6	87	US-09-741-669-441	Sequence 441, App
5	29	80.6	711	US-09-815-242-11203	Sequence 11203, A
6	28	77.8	219	US-09-903-456-38	Sequence 38, Appl
7	28	77.8	280	US-09-903-456-28	Sequence 28, Appl
8	28	77.8	286	US-09-903-456-59	Sequence 59, Appl
9	28	77.8	289	US-09-903-456-71	Sequence 21, Appl
10	28	77.8	289	US-09-903-456-34	Sequence 34, Appl
11	28	77.8	291	US-09-903-456-36	Sequence 36, Appl
12	28	77.8	292	US-09-738-626-6688	Sequence 6688, Ap
13	28	77.8	293	US-09-903-456-19	Sequence 19, Appl
14	28	77.8	301	US-09-903-456-33	Sequence 33, Appl
15	28	77.8	317	US-09-903-456-13	Sequence 13, Appl
16	28	77.8	318	US-09-903-456-25	Sequence 25, Appl
17	28	77.8	2053	US-10-017-216-2	Sequence 2, Appl1
18	27	75.0	66	US-09-816-989A-4	Sequence 4, Appl1
19	27	75.0	76	US-09-963-959-10	Sequence 10, Appl

20	27	75.0	77	9	US-09-963-959-5	Sequence 5, Appl1
21	27	75.0	85	10	US-09-815-242-12315	Sequence 12315, A
22	27	75.0	283	10	US-09-815-242-11150	Sequence 11150, A
23	27	75.0	289	10	US-09-815-242-4948	Sequence 4948, Ap
24	27	75.0	290	10	US-09-815-242-5743	Sequence 5743, Ap
25	27	75.0	293	10	US-09-815-242-10488	Sequence 10488, A
26	27	75.0	293	10	US-09-815-242-12742	Sequence 12742, A
27	27	75.0	316	9	US-09-895-913A-214	Sequence 214, App
28	27	75.0	346	10	US-09-815-242-13437	Sequence 13437, A
29	27	75.0	376	10	US-09-815-242-13636	Sequence 13636, A
30	27	75.0	341	10	US-09-815-242-10777	Sequence 10777, A
31	27	75.0	404	10	US-09-764-864-1244	Sequence 1244, Ap
32	27	75.0	451	9	US-09-738-626-5058	Sequence 5058, Ap
33	27	75.0	655	10	US-09-805-658-57	Sequence 57, Appl
34	27	75.0	655	10	US-09-844-353A-57	Sequence 57, Appl
35	26	72.2	105	9	US-10-000-235A-162	Sequence 162, App
36	26	72.2	183	10	US-09-731-872-420	Sequence 420, App
37	26	72.2	189	9	US-09-992-738-1	Sequence 1, Appl1
38	26	72.2	205	9	US-10-042-141-103	Sequence 103, App
39	26	72.2	205	10	US-09-726-643-103	Sequence 103, App
40	26	72.2	218	9	US-09-966-614-4	Sequence 4, Appl1
41	26	72.2	251	10	US-09-764-864-860	Sequence 860, App
42	26	72.2	259	10	US-09-764-864-1313	Sequence 1313, Ap
43	26	72.2	316	10	US-09-881-752A-122	Sequence 122, App
44	26	72.2	320	9	US-09-996-634-139	Sequence 139, App
45	26	72.2	320	9	US-09-997-181-139	Sequence 139, App

ALIGNMENTS

RESULT 1
US-09-765-086-202
; Sequence 202, Application US/09765086
; Patent No. US20010046498A1
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Pasqualini, Renata
; APPLICANT: Madh, Arap
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Ellerby, H. Michael
; TITLE OF INVENTION: Chimeric Prostate-Homing Peptides With
; FILE REFERENCE: P-LI 3844
; CURRENT APPLICATION NUMBER: US/09/765, 086
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US 09/489,582
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 235
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-09-765-086-202

Query Match 83.3%; Score 30; DB 10; Length 14;
Best Local Similarity 87.5%; Pred. No. 3.8;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKAAKRAA 8
DB 7 AKAAKRAA 14

RESULT 2
US-09-815-242-5166
; Sequence 5166, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: Prokaryotes
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5166
LENGTH: 101
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5166

Query Match
Best Local Similarity 83.3%; Score 30; DB 10; Length 101;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
111111
DB 18 AKYAKKRA 25

RESULT 3
US-09-908-322-10
Sequence 10, Application US/09908322
Patent No. US20020107194A1
GENERAL INFORMATION:
APPLICANT: Ish-Horowitz, David
Henrique, Domingos Manuel Pinto
Lewis, Julian Hart
Artavanis-Tsakonas, Spyridon
Gray, Grace
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF
VERTEBRATE DELTA GENE AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/908,322
FILING DATE: 17-Jul-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/981,392
FILING DATE: 22-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-123
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-908-322-10

Query Match
Best Local Similarity 80.6%; Score 29; DB 10; Length 45;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAKKR 7
111111
DB 21 AKAKKR 27

RESULT 4
US-09-741-669-441
Sequence 441, Application US/09741669
Patent No. US20020022718A1
GENERAL INFORMATION:
APPLICANT: Forsyth, R. Allyn
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
TITLE OF INVENTION: Genes Identified as required for
FILE REFERENCE: ELITRA.009A
CURRENT APPLICATION NUMBER: US/09/741,669
CURRENT FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: 60/173005
PRIOR FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 441
LENGTH: 87
TYPE: PRT
ORGANISM: Escherichia coli
US-09-741-669-441

Query Match
Best Local Similarity 80.6%; Score 29; DB 10; Length 87;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KAARKRA 8
111111
DB 5 KSAKKRA 11

RESULT 5
US-09-815-242-11203
Sequence 11203, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Yamamoto, Robert T.

```
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11203
LENGTH: 711
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-815-242-11203
```

```
Query Match      80.6%; Score 29; DB 10; Length 711;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AKAARR 7
Db      399 ARAAKR 405
```

```
RESULT 6
US-09-903-456-38
Sequence 38, Application US/09903456
Patent No. US20020138874A1
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Mukerji, Pradip
APPLICANT: Leonard, Amanda Eun-Yeong
APPLICANT: Huang, Yung-Sheng
APPLICANT: Pereira, Suzette L.
TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
FILE REFERENCE: 6407.US.P3
CURRENT APPLICATION NUMBER: US/09/903,456
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: US 09/624,670
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: US 09/379,095
PRIOR FILING DATE: 1999-08-23
PRIOR APPLICATION NUMBER: US 09/145,828
PRIOR FILING DATE: 1998-09-02
NUMBER OF SEQ ID NOS: 116
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 38
LENGTH: 219
TYPE: PRT
ORGANISM: Mortierella alpina
US-09-903-456-38
```

```
Query Match      77.8%; Score 28; DB 10; Length 219;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AKAARR 7
Db      189 AKAARR 195
```

```
RESULT 7
US-09-903-456-28
Sequence 28, Application US/09903456
Patent No. US20020138874A1
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Mukerji, Pradip
APPLICANT: Leonard, Amanda Eun-Yeong
APPLICANT: Huang, Yung-Sheng
APPLICANT: Pereira, Suzette L.
TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
FILE REFERENCE: 6407.US.P3
CURRENT APPLICATION NUMBER: US/09/903,456
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: US 09/624,670
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: US 09/379,095
PRIOR FILING DATE: 1999-08-23
PRIOR APPLICATION NUMBER: US 09/145,828
PRIOR FILING DATE: 1998-09-02
NUMBER OF SEQ ID NOS: 116
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 280
TYPE: PRT
ORGANISM: Mortierella alpina
FEATURE:
NAME/KEY: VARIANT
LOCATION: (280)...(280)
OTHER INFORMATION: Xaa = Unknown or Other at position 280
US-09-903-456-28
```

```
Query Match      77.8%; Score 28; DB 10; Length 280;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AKAARR 7
Db      249 AKAARR 255
```

```
RESULT 8
US-09-903-456-59
Sequence 59, Application US/09903456
Patent No. US20020138874A1
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories
APPLICANT: Mukerji, Pradip
APPLICANT: Leonard, Amanda Eun-Yeong
APPLICANT: Huang, Yung-Sheng
APPLICANT: Pereira, Suzette L.
TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
FILE REFERENCE: 6407.US.P3
CURRENT APPLICATION NUMBER: US/09/903,456
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: US 09/624,670
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: US 09/379,095
PRIOR FILING DATE: 1999-08-23
PRIOR APPLICATION NUMBER: US 09/145,828
PRIOR FILING DATE: 1998-09-02
NUMBER OF SEQ ID NOS: 116
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 59
LENGTH: 286
TYPE: PRT
ORGANISM: Mortierella alpina
US-09-903-456-59
```

```
Query Match      77.8%; Score 28; DB 10; Length 286;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 AKAARKR 7
|||||:1
DB 256 AKAARKR 262

RESULT 9

US-09-903-456-21
; Sequence 21, Application US/09903456
; Patent No. US20020138874A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Mukerji, Pradip
; APPLICANT: Leonard, Amanda Eun-Yeong
; APPLICANT: Huang, Yung-Sheng
; APPLICANT: Pereira, Suzette L.
; TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
; FILE REFERENCE: 6407 US.P3
; CURRENT APPLICATION NUMBER: US/09/903,456
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US 09/624,670
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: US 09/379,095
; PRIOR FILING DATE: 1999-08-23
; PRIOR APPLICATION NUMBER: US 09/145,828
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Mortierella alpina
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (289) .. (289)
; OTHER INFORMATION: Xaa = Unknown or other at position 289
US-09-903-456-21

Query Match 77.8%; Score 28; DB 10; Length 289;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKAARKR 7
|||||:1
DB 258 AKAARKR 264

RESULT 10

US-09-903-456-34
; Sequence 34, Application US/09903456
; Patent No. US20020138874A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Mukerji, Pradip
; APPLICANT: Leonard, Amanda Eun-Yeong
; APPLICANT: Huang, Yung-Sheng
; APPLICANT: Pereira, Suzette L.
; TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
; FILE REFERENCE: 6407 US.P3
; CURRENT APPLICATION NUMBER: US/09/903,456
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US 09/624,670
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: US 09/379,095
; PRIOR FILING DATE: 1999-08-23
; PRIOR APPLICATION NUMBER: US 09/145,828
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Mortierella alpina

US-09-903-456-34

Query Match 77.8%; Score 28; DB 10; Length 289;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKAARKR 7
|||||:1
DB 259 AKAARKR 265

RESULT 11

US-09-903-456-36
; Sequence 36, Application US/09903456
; Patent No. US20020138874A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Mukerji, Pradip
; APPLICANT: Leonard, Amanda Eun-Yeong
; APPLICANT: Huang, Yung-Sheng
; APPLICANT: Pereira, Suzette L.
; TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
; FILE REFERENCE: 6407 US.P3
; CURRENT APPLICATION NUMBER: US/09/903,456
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US 09/624,670
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: US 09/379,095
; PRIOR FILING DATE: 1999-08-23
; PRIOR APPLICATION NUMBER: US 09/145,828
; PRIOR FILING DATE: 1998-09-02
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Mortierella alpina
US-09-903-456-36

Query Match 77.8%; Score 28; DB 10; Length 291;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKAARKR 7
|||||:1
DB 261 AKAARKR 267

RESULT 12

US-09-738-626-6688
; Sequence 6688, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KETKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: PatentIn ver. 3.0
 SEQ ID NO 6688
 LENGTH: 292
 TYPE: PRT
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6688

Query Match 77.8%; Score 28; DB 9; Length 292;
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARR 7
 Db 276 AKVAKR 282

RESULT 13
 US-09-903-456-19
 Sequence 19, Application US/09903456
 Patent No. US20020138874A1
 GENERAL INFORMATION:
 APPLICANT: Abbott Laboratories
 APPLICANT: Mukerji, Pradip
 APPLICANT: Leonard, Amanda Eun-Yeong
 APPLICANT: Huang, Yung-Sheng
 APPLICANT: Pereira, Suzette L.
 TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
 FILE REFERENCE: 6407 US P3
 CURRENT APPLICATION NUMBER: US/09/903,456
 PRIOR FILING DATE: 2001-07-11
 PRIOR APPLICATION NUMBER: US 09/624,670
 PRIOR FILING DATE: 2000-07-24
 PRIOR APPLICATION NUMBER: US 09/379,095
 PRIOR FILING DATE: 1999-08-23
 PRIOR APPLICATION NUMBER: US 09/145,828
 PRIOR FILING DATE: 1998-09-02
 NUMBER OF SEQ ID NOS: 116
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 19
 LENGTH: 293
 TYPE: PRT
 ORGANISM: Mortierella alpina
 FEATURE:
 NAME/KEY: VARIANT
 LOCATION: (293)...(293)
 OTHER INFORMATION: Xaa - Unknown or Other at position 293
 US-09-903-456-19

Query Match 77.8%; Score 28; DB 10; Length 293;
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARR 7
 Db 262 AKAARR 268

RESULT 14
 US-09-903-456-33
 Sequence 33, Application US/09903456
 Patent No. US20020138874A1
 GENERAL INFORMATION:
 APPLICANT: Abbott Laboratories
 APPLICANT: Mukerji, Pradip
 APPLICANT: Leonard, Amanda Eun-Yeong
 APPLICANT: Huang, Yung-Sheng
 APPLICANT: Pereira, Suzette L.
 TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
 FILE REFERENCE: 6407 US P3
 CURRENT APPLICATION NUMBER: US/09/903,456
 PRIOR FILING DATE: 2001-07-11
 PRIOR APPLICATION NUMBER: US 09/624,670

PRIOR FILING DATE: 2000-07-24
 PRIOR APPLICATION NUMBER: US 09/379,095
 PRIOR FILING DATE: 1999-08-23
 PRIOR APPLICATION NUMBER: US 09/145,828
 PRIOR FILING DATE: 1998-09-02
 NUMBER OF SEQ ID NOS: 116
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 33
 LENGTH: 301
 TYPE: PRT
 ORGANISM: Mortierella alpina
 FEATURE:
 NAME/KEY: VARIANT
 LOCATION: (301)...(301)
 OTHER INFORMATION: Xaa - Unknown or Other at position 301
 US-09-903-456-33

Query Match 77.8%; Score 28; DB 10; Length 301;
 Best Local Similarity 85.7%; Pred. No. 2e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARR 7
 Db 270 AKAARR 276

RESULT 15
 US-09-903-456-13
 Sequence 13, Application US/09903456
 Patent No. US20020138874A1
 GENERAL INFORMATION:
 APPLICANT: Abbott Laboratories
 APPLICANT: Mukerji, Pradip
 APPLICANT: Leonard, Amanda Eun-Yeong
 APPLICANT: Huang, Yung-Sheng
 APPLICANT: Pereira, Suzette L.
 TITLE OF INVENTION: ELONGASE GENES AND USES THEREOF
 FILE REFERENCE: 6407 US P3
 CURRENT APPLICATION NUMBER: US/09/903,456
 PRIOR FILING DATE: 2001-07-11
 PRIOR APPLICATION NUMBER: US 09/624,670
 PRIOR FILING DATE: 2000-07-24
 PRIOR APPLICATION NUMBER: US 09/379,095
 PRIOR FILING DATE: 1999-08-23
 PRIOR APPLICATION NUMBER: US 09/145,828
 PRIOR FILING DATE: 1998-09-02
 NUMBER OF SEQ ID NOS: 116
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13
 LENGTH: 317
 TYPE: PRT
 ORGANISM: Mortierella alpina
 US-09-903-456-13

Query Match 77.8%; Score 28; DB 10; Length 317;
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARR 7
 Db 287 AKAARR 293

Search completed: April 8, 2003, 11:52:45
 Job time : 12.0345 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:18:53 ; Search time 12.6897 Seconds
(without alignments)
60.606 Million cell updates/sec

Title: US-09-496-391-15

Perfect score: 36

Sequence: 1 AKAKKRA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR3:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	91.7	731	2 T01302	hypothetical prote
2	33	91.7	788	2 S70079	1,4-alpha-glucan b
3	32	88.9	326	2 G96812	protein F3F9.6 (Im
4	32	88.9	666	2 S75289	ribonuclease II -
5	31	86.1	168	2 PNU0680	nitrogenase EC 1.
6	31	86.1	499	1 B29042	nitrogenase (EC 1.
7	31	86.1	820	2 T04227	hypothetical prote
8	30	83.3	101	2 H83114	30S ribosomal prot
9	30	83.3	227	2 B81054	deda protein, prob
10	30	83.3	515	2 F70786	probable pepa - My
11	30	83.3	697	2 T16306	hypothetical prote
12	29	80.6	64	2 S70831	ribosomal protein
13	29	80.6	86	2 A82293	ribosomal protein
14	29	80.6	87	1 R3C20	ribosomal protein
15	29	80.6	87	2 A85483	30S ribosomal subu
16	29	80.6	87	2 AEO058	30S ribosomal prot
17	29	80.6	87	2 AEO507	30S ribosomal prot
18	29	80.6	87	2 B90632	30S ribosomal subu
19	29	80.6	132	2 F82246	conserved hypothet
20	29	80.6	133	1 S17727	H+-transporting tw
21	29	80.6	166	2 B55316	NiTX nitrogen fixa
22	29	80.6	185	2 A32137	histone H1-delta -
23	29	80.6	189	2 S44639	hypothetical prote
24	29	80.6	238	2 T02118	hypothetical prote
25	29	80.6	261	2 H75619	hemin ABC transpor
26	29	80.6	281	2 D98307	hypothetical prote
27	29	80.6	281	2 A62975	2-haloalkanoic aci
28	29	80.6	299	2 A95293	Trm17a probable tr
29	29	80.6	347	2 E83525	TO1A protein PA097

30	29	80.6	374	2 T40158	hypothetical prote
31	29	80.6	402	2 S42367	lag-2 protein - Ca
32	29	80.6	423	2 F82295	ATP-dependent RNA
33	29	80.6	433	2 S25194	zuotin - yeast (Sa
34	29	80.6	465	2 D84427	hypothetical prote
35	29	80.6	529	2 T20121	hypothetical prote
36	29	80.6	632	2 A25784	hypothetical 70K p
37	29	80.6	711	2 B64169	phosphate acetyltr
38	29	80.6	739	2 T15215	hypothetical prote
39	29	80.6	1019	2 T00117	dve protein - frui
40	28	77.8	86	2 A01993	hypothetical prote
41	28	77.8	101	2 AB1189	PTS system, IIB co
42	28	77.8	101	2 AB1547	PTS system, IIB co
43	28	77.8	115	2 F70931	hypothetical prote
44	28	77.8	116	2 T01231	hypothetical prote
45	28	77.8	140	2 C95379	hypothetical prote

ALIGNMENTS

RESULT 1
T01302
hypothetical protein T14P8.8 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 14-May-1999
C:Accession: T01302
R:Kalicki, J.; Elliott, G.; Cloud, J.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of A. thaliana T14P8.
A:Reference number: Z14290
A:Accession: T01302
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-731 <KAL>
A:Cross-references: EMBL:AF069298; NID:g3193282; PID:g3193293
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 46/3; 124/1; 154/3; 192/3; 254/1; 469/2; 511/1; 541/3; 625/1; 646/2; 696/1
A:Note: T14P8.8

Query Match
Best Local Similarity 91.7%; Score 33; DB 2; Length 731;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKAKKRA 8
DB 133 SKAKKRA 140
:|||||
|

RESULT 2
S70079
1,4-alpha-glucan branching enzyme (EC 2.4.1.18) I - Streptomyces coelicolor
N:Alternate names: glycogen branching enzyme
C:Species: Streptomyces coelicolor
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 11-Jan-2000
C:Accession: S70079; T42040
R:Bruton, C.J.; Plaskitt, K.A.; Chater, K.F.
Mol. Microbiol. 18, 89-99, 1995
A:Title: Tissue-specific glycogen branching isoenzymes in a multicellular prokaryote,
A:Reference number: S70078; MUID:96154943; PMID:8596463
A:Accession: S70079
A:Molecule type: DNA
A:Residues: 1-788 <BRU>
A:Cross-references: EMBL:X83397; NID:g1061285; PIDN:CAA58314.1; PID:g1061286
A:Experimental source: strain A3(2)
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1993
A:Note: only a part of the nucleic acid sequence is shown
C:Genetics:
A:Gene: glgBI
A:Start codon: GTG

C:Superfamily: 1.4-alpha-glucan branching enzyme
C:Keywords: glycogen/starch biosynthesis; glycosyltransferase; hexosyltransferase

Query Match 91.7%; Score 33; DB 2; Length 788;
Best Local Similarity 87.5%; Pred. No. 56;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
|||||:1
DB 31 AKAKKRA 38

RESULT 3

G96812

protein F3F9.6 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: G96812

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chlin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

anssen, N.F.; Hughes, B.; Hultzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G96812

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-326 <STO>

A:Cross-references: GB:AE005173; NID:g8052548; PIDN:AAF71812.1; GSPDB:GN00141

C:Genetics:

A:Gene: F3F9.6

A:Map position: 1

Query Match

Best Local Similarity 88.9%; Score 32; DB 2; Length 326;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
|||||:1
DB 96 AEAKKRA 103

RESULT 4

S75289

ribonuclease II - Synecocystis sp. (strain PCC 6803)

N:Alternate names: protein sll1290

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S75289

R:Taneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wade, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

S.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S75289

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-666 <RAN>

A:Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BA17203.1; PID:d101793

C:Genetics:

A:Gene: rnb

A:Start codon: GTG

Query Match 88.9%; Score 32; DB 2; Length 666;
Best Local Similarity 87.5%; Pred. No. 77;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
|||||:1
DB 407 AEAKKRA 414

RESULT 5

PN0680

nitrogenase (EC 1.18.6.1) molybdenum-iron protein alpha chain - Rhodobacter capsulatu

C:Species: Rhodobacter capsulatus

C:Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 19-Jan-2001

C:Accession: PN0680; S18918; S22169

R:Willison, J.C.; Pierrat, J.; Huebner, P.

Gene 133, 39-46, 1993

A:Title: Sequence and transcript analysis of the nitrogenase structural gene operon (

A:Reference number: JN0887; MUID:94040794; PMID:7693551

A:Accession: PN0680

A:Molecule type: DNA

A:Residues: 1-168 <MID>

R:Willison, J.C.; Pierrat, J.; Huebner, P.; Chabert, J.; Vignais, P.M.

submitted to the EMBL Data Library, November 1991

A:Description: Northern blot analysis of the nitrogenase structural gene operon (nifH

A:Reference number: S18916

A:Accession: S18918

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 61-168 <MID>

A:Cross-references: EMBL:X63353; NID:g46070; PIDN:CAA44956.1; PID:g46016

A:Accession: S22169

A:Molecule type: DNA

A:Residues: 61-168 <MID>

A:Cross-references: EMBL:X63353; NID:g46070; PIDN:CAA44956.1; PID:g46071

C:Genetics:

A:Gene: nifD

C:Superfamily: dinitrogenase alpha chain; nitrogenase vanadium-iron protein alpha cha

C:Keywords: ATP; iron; iron-sulfur protein; metalloprotein; molybdenum; nitrogen fixa

F;123/Binding site: homocitryl Mo-7Fe-8S cluster molybdenum (His) (ligand) #status pr

OY 1 AKAKKRA 8
|||||:1
DB 25 AKAKKRA 32

RESULT 6

B29042

nitrogenase (EC 1.18.6.1) molybdenum-iron protein alpha chain - Rhodobacter capsulatu

N:Alternate names: dinitrogenase alpha chain; nitrogenase component I alpha chain

C:Species: Rhodobacter capsulatus

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: B29042

R:Schumann, J.P.; Waitches, G.M.; Scolnik, P.A.

Gene 48, 81-92, 1986

A:Title: A DNA fragment hybridizing to a nif probe in Rhodobacter capsulatus is homol

A:Reference number: A91562; MUID:87163519; PMID:3557130

A:Accession: B29042

A:Molecule type: DNA

A:Residues: 1-499 <SCH>

A:Cross-references: GB:M15270; NID:g151971; PIDN:AAA26141.1; PID:g151973

C:Genetics:

A:Gene: nifD

C:Superfamily: dinitrogenase alpha chain; nitrogenase vanadium-iron protein alpha cha

C:Keywords: ATP; iron-sulfur protein; metalloprotein; molybdenum; nitrogen fixation;

F;19-496/Domain: nitrogenase vanadium-iron protein alpha chain homology <VIA>

F;290/Binding site: homocitryl Mo-7Fe-8S cluster (Cys) (covalent) #status predicted

F;454/Binding site: homocitryl Mo-7Fe-8S cluster molybdenum (His) (ligand) #status pr

Query Match 86.1%; Score 31; DB 1; Length 499;
Best Local Similarity 87.5%; Pred. No. 96;

Matches 7: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARKRA 8
||| ||||
Db 29 AKAKKRA 36

RESULT 7
T04227
hypothetical protein FL14M19.10 - Arabidopsis thaliana (fragment)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
C:Accession: T04227
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Hohensei, J.; Mew
submitted to the Protein Sequence Database, March 1999
A:Reference number: 215262
A:Accession: T04227
A:Molecule type: DNA
A:Residues: 1-820 <BEV>
A:Cross-references: EMBL:AL049480
A:Experimental source: cultivar Columbia; BAC clone FL14M19
C:Genetics:
A:Map position: 4
A:Introns: 158/3; 227/3; 243/2; 295/3; 316/2; 379/3; 409/3; 456/2; 495/3
A>Note: FL14M19.10

Query Match 86.1%; Score 31; DB 2; Length 820;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARKRA 8
||| ||||
Db 705 AKARKRA 712

RESULT 8
H83114
30S ribosomal protein S14 PA4250 [Imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83114
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
L.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
gen
A:Reference number: A82950; MUID:20437337; PMID:10954043
A:Accession: H83114
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-101 <STO>
A:Cross-references: GB:AE004841; GB:AE004091; NID:g9950451; PIDN:AAG07638.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: rpsN; PA4250
C:Superfamily: Escherichia coli ribosomal protein S14

Query Match 83.3%; Score 30; DB 2; Length 101;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 7: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARKRA 8
||| ||||
Db 18 AKYAKRA 25

RESULT 9
B81054
deda protein, probable NMB1689 [Imported] - Neisseria meningitidis (strain MC58 serogrou
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: B81054
R:Teitel, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizsa, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: B81054
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-227 <TEM>
A:Cross-references: GB:AE002519; GB:AE002098; NID:g7226937; PIDN:AAE42037.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1689
C:Superfamily: probable alkaline phosphatase yncC

Query Match 83.3%; Score 30; DB 2; Length 227;
Best Local Similarity 87.5%; Pred. No. 79;
Matches 7: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARKRA 8
||| ||||
Db 216 AKAKKAA 223

RESULT 10
F70786
probable pepA - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70786
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Kajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
me
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70786
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-515 <CO>
A:Cross-references: GB:270283; GB:AL123456; NID:g3261561; PIDN:CAA94271.1; PID:g32615
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: pepB
C:Superfamily: cytosol aminopeptidase

Query Match 83.3%; Score 30; DB 2; Length 515;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 7: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAARKRA 8
||| ||||
Db 174 AKDAKRA 181

RESULT 11
T16306
hypothetical protein F40F4.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16306
R:Wilson, R.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F40F4.
A:Reference number: Z18493
A:Accession: T16306
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-697 <WIL>
A:Cross-references: EMBL:U40420; NID:g1065513; PID:g1065515; PIDN:AAA81431.1; CESP:F4
C:Genetics:

A:Gene: CESP:F40F4.7
A:Introns: 14/2; 86/3; 135/3; 223/1; 244/1; 331/1; 365/1; 396/1; 435/1; 559/3; 627/2; 65

Query Match 83.3%; Score 30; DB 2; Length 697;

Best Local Similarity 75.0%; Pred. No. 2e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKAKKRA 8
DB 467 KSAKKRA 474

RESULT 12

S70831

ribosomal protein S21 - Myxococcus xanthus

C:Species: Myxococcus xanthus

C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 13-Aug-1999

C:Accession: S70831

R:Davis, J.M.; Mayor, J.; Plamann, L.

Mol. Microbiol. 18, 943-952, 1995

A:Title: A missense mutation in rpoD results in an A-signalling defect in Myxococcus xan

A:Reference number: S70829; MUID:96422481; PMID:8825098

A:Accession: S70831

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-64 <DAV>

A:Cross-references: EMBL:U20669; NID:9710339; PIDN:AAH60205.1; PID:9710340

C:Genetics:

A:Gene: rpsu

C:Superfamily: Escherichia coli ribosomal protein S21

C:Keywords: protein biosynthesis; ribosome

Query Match 80.6%; Score 29; DB 2; Length 64;

Best Local Similarity 87.5%; Pred. No. 43;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKAKKRA 8
DB 49 ALAKKRA 56

RESULT 13

AB82293

ribosomal protein S20 VC0679 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: AB82293

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chardson, D.; Ermoaleva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: AB82035; MUID:20406833; PMID:10952301

A:Accession: AB82293

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-86 <HEI>

A:Cross-references: GB:AE004154; GB:AE003852; NID:99655115; PIDN:AAF93844.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0679

A:Map position: 1

C:Superfamily: Escherichia coli ribosomal protein S20

Query Match 80.6%; Score 29; DB 2; Length 86;

Best Local Similarity 85.7%; Pred. No. 56;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KAARKRA 8
DB 5 KSAKKRA 11

RESULT 14

R3BC20

ribosomal protein S20/L26 [validated] - Escherichia coli (strain K-12)

N:Alternate names: ribosomal protein L26; ribosomal protein S20

C:Species: Escherichia coli

C:Date: 15-Oct-1982 #sequence_revision 15-Oct-1982 #text_change 01-Mar-2002

C:Accession: A30425; A02748; S40547; G64722; S07374

R:Mackie, G.A.

J. Biol. Chem. 256, 8177-8182, 1981

A:Title: Nucleotide sequence of the gene for ribosomal protein S20 and its flanking r

A:Reference number: A30425; MUID:81264207; PMID:6267039

A:Accession: A30425

A:Molecule type: DNA

A:Residues: 1-87 <MAC>

A:Cross-references: GB:D10483; GB:J01597; GB:J01683; GB:J01706; GB:K01298; GB:K01990;

IDN:BAA01302.1; PID:9285753; GB:X04382; GB:V00345; NID:942864; PIDN:CAA27968.1; PID:9

RTM:Altman-liebold, B.; Marzinzig, E.; Lehmann, A.

FEBS Lett. 68, 110-114, 1976

A:Title: Primary structure of protein S20 from the small ribosomal subunit of Escheri

A:Reference number: A02748; MUID:77003692; PMID:786731

A:Accession: A02748

A:Molecule type: protein

A:Residues: 2-87 <WTT>

A:Experimental source: strain K

R:Tura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Miz

submitted to the EMBL Data Library, December 1992

A:Description: Systematic sequencing of the Escherichia coli genome: analysis of the

A:Reference number: S40531

A:Accession: S40547

A:Molecule type: DNA

A:Residues: 1-2-87 <YUR>

A:Cross-references: EMBL:D10483; NID:9216434; PIDN:BAA01302.1; PID:9285753

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64722

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-87 <BLAT>

A:Cross-references: GB:AE000113; GB:U00096; NID:92367095; PIDN:AACT3134.1; PID:917862

A:Experimental source: strain K-12, substrain MG1655

R:Arnold, R.J.; Reilly, J.P.

Anal. Biochem. 269, 105-112, 1999

A:Title: Observation of Escherichia coli ribosomal proteins and their posttranslation

A:Reference number: A59071; MUID:99196679; PMID:10094780

A:Contents: annotation; mass spectrographic analysis

A:Note: mass spectrographic analysis of post-translational modifications; any acid la

C:Genetics:

A:Gene: rps2

A:Map position: 0 min

A:Start codon: GTG

C:Complex: the ribosome is composed of the large (50S) and small (30S) subunit; the 1

S rRNA and 22 distinct proteins

C:Complex: small subunit ribosomal proteins: S1 (PIR:R3EC1), S2 (PIR:R3EC2), S3 (PIR:

R3EC10), S11 (PIR:R3EC11), S12 (PIR:R3EC12), S13 (PIR:R3EC13), S14 (PIR:R3EC14), S15

IR:R3EC21), S22 (PIR:C64901) [validated; MUID:99196679]

C:Function:

A:Pathway: protein biosynthesis

A:Note: binds 16S rRNA

C:Superfamily: Escherichia coli ribosomal protein S20

C:Keywords: protein biosynthesis; ribosome; RNA binding

F:2-87/Product: ribosomal protein S20 #status experimental <MAT>

Query Match 80.6%; Score 29; DB 1; Length 87;

Best Local Similarity 85.7%; Pred. No. 56;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KAARKRA 8
DB 5 KSAKKRA 11

RESULT 15

A85483

30S ribosomal subunit protein S20 [imported] - Escherichia coli (strain O157:H7, substr

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002

C:Accession: A85483

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: A85483

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-87 <STO>

A:Cross-references: GB:AE005174; NID:q12512706; PIDN:AG54325.1; GSPDB:GN00145; UWGP:Z00

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: rpsT

C:Superfamily: Escherichia coli ribosomal protein S20

Query Match

80.6%; Score 29; DB 2; Length 87;

Best Local Similarity 85.7%; Pred. No. 56;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KAARKRA 8

1:|||||

Db 5 KSAKKRA 11

Search completed: April 8, 2003, 11:30:01
Job time : 15.6897 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:12:02 ; Search time 6.2069 Seconds

(without alignments)
53.458 Million cell updates/sec

Title: US-09-496-391-15

Perfect score: 36

Sequence: 1 AKAARRA 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	32	88.9	480	1 NIFD_CYAAS	007642 cyanothec
2	32	88.9	666	1 RN2H_SYNY3	P73177 synechocyst
3	31	86.1	499	1 NIFD_RHOCA	P08717 rhodobacter
4	30	83.3	101	1 RS14_PSEAE	Q9hwe8 pseudomonas
5	30	83.3	515	1 AMPA_MVCTU	Q10401 mycobacteri
6	29	80.6	64	1 RS21_MYXXA	P49225 myxococcus
7	29	80.6	86	1 RS20_ECOLI	P02378 escherichia
8	29	80.6	86	1 RS20_HAEIN	P44959 haemophilus
9	29	80.6	86	1 RS20_SALTU	P41787 salmonella
10	29	80.6	87	1 RS20_VIRCH	O34239 vibrio chol
11	29	80.6	87	1 RS20_VIRCH	O8zinz versinia pe
12	29	80.6	89	1 RS20_PASMU	Q9ckx0 pasteurella
13	29	80.6	133	1 ATPE_BACFI	P22480 bacillus fi
14	29	80.6	185	1 H1D_STRPU	P15870 strongyloce
15	29	80.6	189	1 YPT2_CAEEL	P50600 pseudomonas
16	29	80.6	347	1 TOLA_PSEAE	P55642 rhizobium s
17	29	80.6	390	1 Y4RI_RHITSN	P45442 caenorhabdi
18	29	80.6	402	1 LAG2_CAEEL	P32527 saccharomyc
19	29	80.6	433	1 ZUO1_YEAST	Q9j142 mus musculi
20	29	80.6	708	1 GIT2_MOUSE	P7A_HAEIN
21	29	80.6	710	1 PTA_HAEIN	Q14161 haemophilus
22	29	80.6	759	1 GIT2_HUMAN	O14161 homo sapien
23	29	80.6	761	1 GIT1_HUMAN	O9y2x7 homo sapien
24	29	80.6	770	1 GIT1_RAT	O9z272 rattus norv
25	28	77.8	138	1 ATPE_THETN	Q8rc14 thermoaer
26	28	77.8	146	1 ATPE_LACAC	Q9rgy0 lactobacill
27	28	77.8	162	1 RL24_HORVU	P50888 hordeum vul
28	28	77.8	194	1 HL_SALTU	P02254 salmo trutt
29	28	77.8	206	1 H1_ONCMV	P06350 oncorhynch
30	28	77.8	213	1 LEAL_HORVU	P14928 hordeum vul
31	28	77.8	247	1 VPM_BPP2	P25476 bacterioph
32	28	77.8	256	1 NIFD_NOSCO	P2337 nostoc comm
33	28	77.8	316	1 TAL_HELPJ	Q9zjcs helicobacte

34	28	77.8	346	1 RS6_AEDAE	Q9u761 aedes aegy
35	28	77.8	358	1 OPED_BACSU	P24136 bacillus su
36	28	77.8	444	1 PUR8_ARCFU	O28041 archaeoglob
37	28	77.8	673	1 FLID_TREMA	Q9kww7 treponema m
38	28	77.8	684	1 RPSD_AGRRT5	P33452 agrobacteri
39	28	77.8	684	1 RPSD_RHIME	Q59753 rhizobium m
40	28	77.8	687	1 TRA_HAEIN	O05069 haemophilus
41	28	77.8	699	1 NP14_HUMAN	Q14978 homo sapien
42	28	77.8	722	1 COAT_PAVIH	P03136 hamster par
43	28	77.8	807	1 YMT9_YEAST	O04660 saccharomyc
44	28	77.8	821	1 RNR_VIRCH	Q9kny1 vibrio chol
45	28	77.8	918	1 DNLI_RAT	Q9jhy8 rattus norv

ALIGNMENTS

RESULT 1	ID	NIFD_CYAAS	STANDARD	PRT	480 AA.
AC	007642				
DT	15-JUL-1999	(Rel. 38, Created)			
DT	15-JUL-1999	(Rel. 38, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Nitrogenase molybdenum-iron protein alpha chain (EC 1.18.6.1)				
DE	(Nitrogenase component I) (Dinitrogenase).				
GN	NIFD.				
OS	Cyanothecae (strain ATCC 51142).				
OC	Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.				
OX	NCBI_TaxID=43989;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20062360; PubMed=10594374;				
RA	Colton-Lopez M.S., Tang H., Tucker D.L., Sherman L.A.;				
RT	"Analysis of the nifHDK operon and structure of the NifH protein from				
RT	the unicellular, diazotrophic cyanobacterium, Cyanothecae strain sp.				
RT	ATCC 51142(1).";				
RL	Biochim. Biophys. Acta 1473:363-375(1999).				
CC	-1- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE				
CC	CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE				
CC	IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.				
CC	-1- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP				
CC	= 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate.				
CC	-1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS THAT BINDS				
CC	30-32 FE. 2 MO. AND INORGANIC SULFUR.				
CC	-1- SIMILARITY: BELONGS TO THE NIFD/NIFE/NIFN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib.ch).				
CC	-----				
DR	EMBL: AF003337; AAB61283.1; -				
DR	HSSP: P07328; 3MIN.				
DR	InterPro: IPR000318; Nitrogenase-comp1.				
DR	InterPro: IPR000510; Oxired_nitrogensel.				
DR	Pfam: PF00148; Oxidored_nitro; 1.				
DR	TIGRFAMS: TIGR01282; nifd; 1.				
DR	PROSITE: PS00699; NITROGENASE_1.1; 1.				
DR	PROSITE: PS00699; NITROGENASE_1.2; 1.				
KW	Oxidoreductase; Nitrogen fixation; Molybdenum; Iron-sulfur.				
SQ	SEQUENCE 480 AA; 53682 MW; 30830095EC405E05 CRC64;				
QY	2 KAARKRA 8				
DB	21 KAARKRA 27				

Query Match 88.9%; Score 32; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 2
ID RN2H_SVNY3 STANDARD: PRT: 666 AA.
AC P73177:
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Hypothetical ribonuclease sl11290 (EC 3.1.1.1).
GN SL11290.
OS Synecocystis sp. (strain RCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nariu K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain RCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90904; BAA17203.1; -
DR InterPro: IPR001900; Ribonuclease-II.
DR Pfam: PF00773; RNB; 1.
DR Hypothetical protein: Hydrolase; Nuclease; Complete proteome.
SQ SEQUENCE 666 AA; 75802 MW; 6DB661D7ADC9A753 CRC64;

Query Match
Best Local Similarity 88.9%; Score 32; DB 1; Length 666;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
DB 407 AKAKKRA 414

RESULT 3
ID NIFD_RHOCA STANDARD: PRT: 499 AA.
AC P08717:
DT 01-JAN-1988 (rel. 06, Created)
DT 01-JAN-1988 (rel. 06, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE Nitrogenase molybdenum-iron protein alpha chain (EC 1.18.6.1)
DE (Nitrogenase component I) (Dinitrogenase).
GN NIFD.
OS Rhodospirillum rubrum (Rhodospirillum rubrum).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
OC Rhodospirillum.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87163519; PubMed=3557130;
RA Schumann J.P., Walchli G.M., Scolnik P.A.;
RT "A DNA fragment hybridizing to a nif probe in Rhodospirillum rubrum
RT is homologous to a 16S rRNA gene.";
RL Gene 48:81-92(1986).
CC -1- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE

```

```

CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE
CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.
CC -1- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP
CC = 8 oxidized ferredoxin + 2 NH(3) + 16 ADP + 16 phosphate.
CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS THAT BINDS
CC 30-32 FE, 2 MO, AND INORGANIC SULFUR.
CC -1- SIMILARITY: BELONGS TO THE NIFD/NIFE/NIFN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15370; AAA26141.1; -
DR PIR: B29042; B29042.
DR HSP: P07328; 3MTN.
DR InterPro: IPR000318; Nitrogenase-comp1.
DR InterPro: IPR000510; Oxred_nitrogenase1.
DR Pfam: PF00148; oxidored_nitro; 1.
DR TIGR: TIGR01282; nifD; 1.
DR PROSITE: PS00090; NITROGENASE_1_2; 1.
DR PROSITE: PS00699; NITROGENASE_1_1; 1.
DR OXIDOREDUCTASE: Nitrogen fixation; Molybdenum, iron-sulfur.
SQ SEQUENCE 499 AA; 56105 MW; C39786673C22C949 CRC64;

Query Match
Best Local Similarity 86.1%; Score 31; DB 1; Length 499;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
DB 29 AKAKKRA 36

RESULT 4
ID RS14_PSEAE STANDARD: PRT: 101 AA.
AC Q9HWE8;
DT 15-JUN-2002 (rel. 41, Created)
DT 15-JUN-2002 (rel. 41, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE 30S ribosomal protein S14.
DE RPSN OR PA4250.
GN GN
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Ladd K., Lam R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RT Nature 406:959-964(2000).
CC -1- FUNCTION: Known to be required for the assembly of 30S particles
CC and may also be responsible for determining the conformation of
CC the 16S rRNA at the A site (by similarity).
CC -1- SIMILARITY: BELONGS TO THE S14 FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE004841; AAC07638.1; -
 DR InterPro: IPR001209; Ribosomal_S14.
 DR Pfam: PF00253; Ribosomal_S14; 1.
 DR PROSITE: PS00527; RIBOSOMAL_S14; 1.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 101 AA; 11565 MW; B9D694FDC554902 CRC64;

Query Match 83.3%; Score 30; DB 1; Length 101;
 Best Local Similarity 87.5%; Pred. No. 12;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAAKRA 8
 |||||||
 Db 18 AKYAKRA 25

RESULT 5
 ID AMPA_MYCTU STANDARD: PRT; 515 AA.
 AC Q10401;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase)
 DE (LAP) (Leucyl aminopeptidase)
 CN PEPA OR PEPB OR RV213 OR MT2269 OR MTCY190.24.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Stulton J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
 CC TURNOVER OF INTRACELLULAR PROTEINS. CATALYZES THE REMOVAL OF
 CC UNSUBSTITUTED N-TERMINAL AMINO ACIDS FROM VARIOUS PEPTIDES (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-
 CC Xbb-, in which Xaa is preferably Leu, but may be other amino acids
 CC including Pro although not Arg or Lys, and Xbb may be Pro.
 CC -1- COFACTOR: MANGANESE (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z70283; CA94271.1; -
 DR EMBL: AE007072; AAK46555.1; -
 DR HSSP: P00727; ILAM.
 DR MEROPS: M17.003; -
 DR TIGR: MT2269; -
 DR TubercuList: RV2213; -
 DR InterPro: IPR00819; Peptidase_M17.
 DR Pfam: PF00883; Peptidase_M17; 1.
 DR Pfam: PF02789; Peptidase_M17_N; 1.
 DR PRINTS: PR00481; LAMNOPTDASE.
 DR PROSITE: PS00631; CYTOSOL_AP; 1.
 KW Hydrolase; Aminopeptidase; Manganese; Complete proteome.
 FT METAL 279 279
 FT METAL 284 284 MANGANESE 2 (POTENTIAL).
 FT METAL 302 302 MANGANESE 1 AND 2 (POTENTIAL).
 FT METAL 361 361 MANGANESE 2 (POTENTIAL).
 FT METAL 363 363 MANGANESE 1 (POTENTIAL).
 FT ACT_SITE 291 291 MANGANESE 1 AND 2 (POTENTIAL).
 FT ACT_SITE 365 365 POTENTIAL.
 SQ SEQUENCE 515 AA; 53481 MW; 0A40D181F35798F CRC64;

Query Match 83.3%; Score 30; DB 1; Length 515;
 Best Local Similarity 87.5%; Pred. No. 53;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAAKRA 8
 |||||||
 Db 174 AKDAKRA 181

RESULT 6
 ID RS21_MYXXA STANDARD: PRT; 64 AA.
 AC P49225;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 30S ribosomal protein S21.
 GN RPSU.
 OS Myxococcus xanthus.
 CC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 CC Myxococcales; Cylobacterineae; Myxococcaceae; Myxococcus.
 OX NCBI_TaxID=34;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DKS060;
 RA Davis J.M., Mayor J., Piamann L.;
 RL Submitted (Feb-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE S21P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U20669; AAB60205.1; -
 DR InterPro: IPR001911; Ribosomal_S21.
 DR Pfam: PF01165; Ribosomal_S21; 1.
 DR PRINTS: PR00976; RIBOSOMAL_S21.
 DR ProDom: PD005521; RIBOSOMAL_S21; 1.
 DR TIGRFAMS: TIGR00030; S21P; 1.
 DR PROSITE: PS01181; RIBOSOMAL_S21; 1.
 KW Ribosomal protein.
 SQ SEQUENCE 64 AA; 7352 MW; CAEC3F7BC4F0F CRC64;

Query Match 80.6%; Score 29; DB 1; Length 64;
 Best Local Similarity 87.5%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
 DB 49 ALAKKRA 56

RESULT 7
 RS20_ECOLI STANDARD; PRT; 86 AA.
 AC P02378;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 30S ribosomal protein S20.
 GN RPS20 OR B0023 OR Z0027 OR ECS0026.
 OS Escherichia coli, and
 OC Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=81264207; PubMed=6267039;
 RA Mackie G.A.;
 RT "Nucleotide sequence of the gene for ribosomal protein S20 and its
 RT flanking regions.";
 RL J. Biol. Chem. 256:8177-8182(1981).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87016337; PubMed=2429258;
 RA Mackie G.A.;
 RT "Structure of the DNA distal to the gene for ribosomal protein S20 in
 RT Escherichia coli K12: presence of a strong terminator and an ISI
 RT element.";
 RL Nucleic Acids Res. 14:6965-6981(1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85182715; PubMed=2985604;
 RA Kamio Y., Lin C.-K., Regue M., Wu H.C.;
 RT "Characterization of the illeS-lsp operon in Escherichia coli.
 RT Identification of an open reading frame upstream of the illeS gene and
 RT potential promoter(s) for the illeS-lsp operon.";
 RL J. Biol. Chem. 260:5616-5620(1985).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=92334977; PubMed=1630901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
 RA Isura K., Mizobuchi K., Nakata A.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 RT the 0-2.4 min region.";
 RL Nucleic Acids Res. 20:3305-3308(1992).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grothbeck E.J., Davis N.W., Lim A., Dinalanta E.T., Potamouzis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Ikeda T., Takami H., Hatori M., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [8]
 RP SEQUENCE.
 RC STRAIN-K;
 RX MEDLINE=77003692; PubMed=786731;
 RA Wittmann-Liebold B., Marzinzig E., Lehmann A.;
 RT "Primary structure of protein S20 from the small ribosomal subunit of
 RT Escherichia coli.";
 RL FEBS Lett. 68:110-114(1976).
 RN [9]
 RP MASS SPECTROMETRY.
 RX MEDLINE=99196679; PubMed=10094780;
 RA Arnold R.J., Rellly J.P.;
 RT "Observation of Escherichia coli ribosomal proteins and their
 RT posttranslational modifications by mass spectrometry.";
 RL Anal. Biochem. 269:105-112(1999).
 CC -1- FUNCTION: Binds directly to 16S ribosomal RNA.
 CC -1- MASS SPECTROMETRY: MW=9553.6; METHOD=MALDI.
 CC -1- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X04382; CAA27968.1; -;
 DR EMBL: M10428; AAA24604.1; -;
 DR EMBL: D10483; BAA01302.1; -;
 DR EMBL: AE000113; AAC73134.1; -;
 DR EMBL: AE005179; AAG54325.1; -;
 DR EMBL: AP002550; BAB33449.1; -;
 DR PIR: A02748; R3EC20.
 DR EC02DBASE: I012.9; 6TH EDITION.
 DR Ecogene: EG10919; rpsT.
 DR InterPro: IPR002583; Ribosomal_S20p.
 DR Pfam: PF01649; Ribosomal_S20p. 1.
 DR ProDom: PD004231; Ribosomal_S20p. 1.
 KW Ribosomal protein; rRNA-binding; Complete proteome.
 FT INIT_MET 0
 FT CONFLICT 37 0
 FT SEQUENCE 86 AA; 9553 MW; 29F5C9F7B015C16 CRC64;
 SQ

Query Match 80.6%; Score 29; DB 1; Length 86;
 Best Local Similarity 85.7%; Pred. No. 17;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KAARKRA 8
 DB 4 KSAKKRA 10

RESULT 8
 RS20_HAEIN STANDARD; PRT; 86 AA.
 AC P44959;

```

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S20.
GN RPS20_H10965
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Giordano A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Furumann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: Binds directly to 16S ribosomal RNA (By similarity).
CC -1- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: U32777; AAC22624.1; ALT_SEQ.
DR TIGR: H10965; -.
DR InterPro: IPR002583; Ribosomal_S20P.
DR Pfam: PF01649; Ribosomal_S20P; 1.
DR ProDom: PD004231; Ribosomal_S20P; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
FT INIT_MET 0
FT SEQUENCE 86 AA; 9507 MW; 2B56D4BA39060487 CRC64;
SQ
Query Match 80.6%; Score 29; DB 1; Length 86;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 KAAKRA 8
Db 4 KSAKRA 10

```

```

RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RT Nature 413:852-856(2001).
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE OF 8-79 FROM N.A.
RC SPECIES-S. typhimurium;
RX MEDLINE=95367591; PubMed=7640306;
RA Nemec A., Haywood-Farmer A., Mackie G.A.;
RT "Conserved amino acid residues in the primary structure of ribosomal
RT protein S20 from selected Gram-negative bacteria."
RT Biochim. Biophys. Acta 1263:154-158(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-S. typh; STRAIN-CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Cornerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltham T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typh CT18."
RL Nature 413:848-852(2001).
CC -1- FUNCTION: Binds directly to 16S ribosomal RNA (By similarity).
CC -1- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: AE008695; AAL19007.1; -.
DR EMBL: U20491; AAB87001.1; -.
DR EMBL: AL627265; CAD01198.1; -.
DR STyGene; SG10549; rps2.
DR InterPro: IPR002583; Ribosomal_S20P.
DR Pfam: PF01649; Ribosomal_S20P; 1.
DR ProDom: PD004231; Ribosomal_S20P; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
FT INIT_MET 0
FT BY SIMILARITY.
FT SEQUENCE 86 AA; 9524 MW; 4F0D70EF0BDB44BD CRC64;
SQ
Query Match 80.6%; Score 29; DB 1; Length 86;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 KAAKRA 8
Db 4 KSAKRA 10

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor O17 / Serotype O1;
RX MEDLINE=98117066; PubMed=945788;
RA Williams S.G., Carmel-Harel O., Manning P.A.;
RT "A functional homolog of Escherichia coli Nhar in Vibrio cholerae.";
RL J. Bacteriol. 180:762-765(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Ehsen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Hart D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettein H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dracoli I., Sellers P.,
RA Donaldald L., Uterback T., Fleischmann R.D., Nieman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -1- FUNCTION: Binds directly to 16S ribosomal RNA (By similarity).
CC -1- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ002395; CA05374.1; -
DR EMBL: AE004154; AAF93844.1; -
DR TIGR: VC0679; -
DR InterPro: IPR002583; Ribosomal_S20p.
DR Pfam: PF01649; Ribosomal_S20p; 1.
DR ProDom: PD004231; Ribosomal_S20p; 1.
DR Ribosomal protein; rRNA-binding; Complete proteome.
KW CONFLICT 20 H -> D (in Ref. 1).
FT SEQUENCE 86 AA; 9510 MW; 0EB1F4435BBA1BC CRC64;
SQ
Query Match 80.6%; Score 29; DB 1; Length 86;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KAAKRA 8
DB 5 KSAKRA 11
RESULT 11
RS20_YERPE STANDARD; PRT; 87 AA.
ID RS20_YERPE
AC Q8ZIM3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S20.
GN RPS20 OR YP00472.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Hoiden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagers K., Karlyshev A.V.,

```

```

RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -1- FUNCTION: Binds directly to 16S ribosomal RNA (By similarity).
CC -1- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ14143; CAC89328.1; -
DR InterPro: IPR002583; Ribosomal_S20p.
DR Pfam: PF01649; Ribosomal_S20p; 1.
DR ProDom: PD004231; Ribosomal_S20p; 1.
DR Ribosomal protein; rRNA-binding; Complete proteome.
KW SEQUENCE 87 AA; 9789 MW; 98B9C2FEDCFC38B6 CRC64;
SQ
Query Match 80.6%; Score 29; DB 1; Length 87;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KAAKRA 8
DB 5 KSAKRA 11

```

```

RESULT 12
RS20_PASMU STANDARD; PRT; 89 AA.
ID RS20_PASMU
AC Q9CKG0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S20.
GN RPS20 OR PM1659.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- FUNCTION: Binds directly to 16S ribosomal RNA (By similarity).
CC -1- SIMILARITY: BELONGS TO THE S20P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006202; AAK03743.1; -
DR InterPro: IPR002583; Ribosomal_S20p.
DR Pfam: PF01649; Ribosomal_S20p; 1.
DR ProDom: PD004231; Ribosomal_S20p; 1.
DR Ribosomal protein; rRNA-binding; Complete proteome.
KW SEQUENCE 89 AA; 9867 MW; 3A07DCC3FA9B546 CRC64;
SQ
Query Match 80.6%; Score 29; DB 1; Length 89;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Oy 2 KAAKRA 8
 Db 7 KSAKRA 13

RESULT 13
 ATPE_BACFI

ID ATPE_BACFI STANDARD; PRT; 133 AA.

AC P22480;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP synthase epsilon chain (EC 3.6.3.14) (ATP synthase F1 sector
 DE epsilon subunit).
 GN ATPC.
 OS Bacillus firmus.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1399;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OF4;
 RX MEDLINE=92017665; PubMed=1833620;
 RA Ivey D.M., Krulwich T.A.;
 RT "Organization and nucleotide sequence of the atp genes encoding the
 RT ATP synthase from alkaliphilic Bacillus firmus OF4.";
 RL Mol. Gen. Genet. 229:292-300(1991).
 RN [2]
 RP REVISIONS.
 RA Hicks D., Krulwich T.A.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Produces ATP from ADP in the presence of a proton
 CC gradient across the membrane.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(in) = ADP + phosphate +
 CC H(+)(out).
 CC -1- SUBUNIT: F-type atpases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: A, B and C.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE EPSILON CHAIN FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF30160; AAC48364.1; -
 DR PIR; S17727; S17727.
 DR HSSP; P00832; IAOI.
 DR InterPro: IPR001469; ATPsynT_DE.
 DR Pfam; PF00401; ATP-synT_DE.1.
 DR Pfam; PF02823; ATP-synT_DE_N.1.
 DR ProDom; PD000944; ATPsynT_DE.1.
 DR TIGRFAMs; TIGR01216; ATP_synT_eps1.1.
 DR HydroLase; ATP synthesis; CF(1); Hydrogen ion transport.
 KW SEQUENCE 133 AA; 14328 MW; B548ED48D80829CD CRC64;

Query Match 80.6%; Score 29; DB 1; Length 133;
 Best Local Similarity 75.0%; Pred. No. 25;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KAAKRA 8
 Db 93 KAAKRA 100

RESULT 14
 HID_STRPU
 ID HID_STRPU STANDARD; PRT; 185 AA.
 AC P15870;
 DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Histone H1-delta.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86247975; PubMed=2898141;
 RA Lieber T., Angerer L.M., Angerer R.C., Childs G.;
 RT "A histone H1 protein in sea urchins is encoded by a poly(A) + mRNA.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:4123-4127(1988).
 CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
 CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; J03807; AAA30055.1; -
 DR PIR; A3137; A32137.
 DR HSSP; P02259; IHST.
 DR InterPro: IPR001386; Histone_H1/H5.
 DR InterPro: IPR003216; Linkerhist_N.
 DR Pfam; PF00538; Linker_histone; 1.
 DR ProDom; PD000373; Linkerhist_N.1.
 DR SMART; SM00526; H15; 1.
 DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
 KW SEQUENCE 185 AA; 19297 MW; DA6F310F8F946E6 CRC64;

Query Match 80.6%; Score 29; DB 1; Length 185;
 Best Local Similarity 85.7%; Pred. No. 34;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 KAAKRA 8
 Db 18 KAAKRA 24

RESULT 15
 YPT2_CAEL
 ID YPT2_CAEL STANDARD; PRT; 189 AA.
 AC P41880;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Hypothetical 21.6 kDa protein F37A4.2 in chromosome III.
 GN F37A4.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodertinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Fulton L., Waterston R.;
 RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC -----
DR EMBL; 000032; AAA50630.1; -
DR WormBep; F37A4.2; CE00710.
KW Hypothetical protein.
SQ SEQUENCE 189 AA; 21602 MW; 8827BCC630A015D4 CRC64;

Query Match 80.6%; Score 29; DB 1; Length 189;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;

QY 1 AKAAKKRA 8
: | | | | |
Db 113 SKTAKKRA 120

Search completed: April 8, 2003, 11:24:43
Job time : 7.2069 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:17:38 ; Search time 28.4138 Seconds
(without alignments)
58.013 Million cell updates/sec

Title: US-09-496-391-15
Perfect score: 36
Sequence: 1 AKAKKRA 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	91.7	731	10	081285 arabidopsis
2	33	91.7	774	16	0911K4 streptomyc
3	33	91.7	788	2	059833 streptomyc
4	33	91.7	1265	10	08RX24 arabidopsis
5	32	88.9	326	10	09M9F9 arabidopsis
6	32	88.9	329	2	09EVN5 pseudomonas
7	32	88.9	542	10	08S5V9 oryza sativ
8	31	86.1	60	2	052693 rhodobacter
9	31	86.1	821	10	08VZJ4 arabidopsis
10	31	86.1	821	10	09SW10 arabidopsis
11	30	83.3	70	16	08XWK6 ralsionia s
12	30	83.3	106	10	08WVT1 narcissus p
13	30	83.3	198	4	09H8H4 homo sapien
14	30	83.3	208	5	09S5X0 diatrophila
15	30	83.3	219	5	020220 caenorhabdi
16	30	83.3	227	16	09JY90 nelsseria m

17	30	83.3	244	10	08W120 zea mays (m
18	30	83.3	449	5	097452 giardia lam
19	30	83.3	467	4	09H9F1 homo sapien
20	30	83.3	650	5	09VDK7 dtrosophila
21	30	83.3	1337	10	09LV73 arabidopsis
22	30	83.3	1428	17	08TR62 methanosarc
23	29	80.6	116	2	092JY7 rhizobium m
24	29	80.6	118	2	093S27 rhizobium t
25	29	80.6	129	12	067577 bean golden
26	29	80.6	129	12	08QR23 tomato chlo
27	29	80.6	129	12	08QMG7 tomato crum
28	29	80.6	130	16	0984P6 rhizobium 1
29	29	80.6	132	16	09KT53 vibrio chol
30	29	80.6	166	16	0922L1 rhizobium m
31	29	80.6	188	10	093V83 oryza sativ
32	29	80.6	191	16	093J85 streptomyc
33	29	80.6	238	10	080683 arabidopsis
34	29	80.6	250	16	092M35 rhizobium m
35	29	80.6	261	16	09R2U5 delinococcus
36	29	80.6	281	16	08UA66 agrobacteri
37	29	80.6	299	16	0930E9 rhizobium m
38	29	80.6	332	10	08SR49 oryza sativ
39	29	80.6	374	3	094280 schizosacch
40	29	80.6	377	5	08SV75 enccephalito
41	29	80.6	423	16	09K063 vibrio chol
42	29	80.6	429	13	09W6U3 fuqu rubrip
43	29	80.6	444	16	0925Z4 rhizobium m
44	29	80.6	456	16	092PS3 rhizobium m
45	29	80.6	458	10	09ZU93 arabidopsis

ALIGNMENTS

RESULT 1

ID	081285	PRELIMINARY:	PRT:	731 AA.
AC	081285;			
DT	01-NOV-1998 (TREMBlrel. 08. Created)			
DT	01-NOV-1998 (TREMBlrel. 08. Last sequence update)			
DT	01-OCT-2000 (TREMBlrel. 15. Last annotation update)			
DE	TI4P8.8 protein (AT4G02480 protein).			
GN	TI4P8.8 OR AT4G02480.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	Eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RA	WASHU;			
RT	"The A. thaliana Genome Sequencing Project.;"			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RA	Kalicki J., Elliott G., Cloud J.;			
RT	"The sequence of A. thaliana TI4P8.;"			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. COLUMBIA;			
RA	Waterston R.;			
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RA	Lamer B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.;			
RL	Mayer K.F.X.;			
RN	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
RP	[5]			
RA	SEQUENCE FROM N.A.			
RC	EU Arabidopsis sequencing project;			

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF069298; AAC19277.1; -
 DR EMBL: AL161494; CAB80741.1; -
 SQ SEQUENCE 731 AA; 77422 MW; 8E3E585F79F15562 CRC64;

Query Match 91.7%; Score 33; DB 10; Length 731;
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKRA 8
 Db 133 SKAARKRA 140

RESULT 2

O9L1K4 PRELIMINARY; PRT; 774 AA.

AC O9L1K4: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE 1,4-alpha-glucan branching enzyme.
 GN GLGB1 OR SC05440 OR SC6A11.16C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Bentley S.D.; Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashl H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 [4]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RT Nature 417:141-147(2002).
 RU EMBL: AL138978; CAB72416.1; -
 DR InterPro: IPR00461; Alpha_amyase.
 DR InterPro: IPR004193; Isoamylase_N.
 DR Pfam: PF00128; alpha-amyase; 1.
 DR Pfam: PF02922; Isoamylase_N; 1.
 SQ SEQUENCE 774 AA; 85450 MW; 1D5555330F249BF CRC64;

Query Match 91.7%; Score 33; DB 16; Length 774;
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKRA 8

Db 31 AKAARKRA 38

RESULT 3

O59833 PRELIMINARY; PRT; 788 AA.

AC O59833: 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) (glycogen branching
 enzyme) (AMYLO-1,4 to 1,6)TRANSGLUCOSIDASE.
 GN GLGB1.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE=96154943; PubMed=8596463;
 RA Bruton C.J., Plaskitt K.A., Chater K.F.;
 RT "Tissue-specific glycogen branching isoenzymes in a multicellular
 prokaryote, Streptomyces coelicolor A3(2).";
 RL Mol. Microbiol. 18:89-99(1995).
 RN RP SEQUENCE OF 1-217 FROM N.A.
 RC STRAIN-A3(2);
 RA Schneider D., Bruton C.J., Chater K.F.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -I CATALYTIC ACTIVITY: FORMATION OF 1,6-GLUCOSIDIC LINKAGES OF
 GLYCOGEN.
 DR EMBL: X83397; CAA58314.1; -
 DR EMBL: AJ001205; CAA04603.1; -
 DR InterPro: IPR00461; Alpha_amyase.
 DR InterPro: IPR004193; Isoamylase_N.
 DR Pfam: PF00128; alpha-amyase; 1.
 DR Pfam: PF02922; Isoamylase_N; 1.
 DR KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 788 AA; 87268 MW; 43DF1A6F08022A67 CRC64;

Query Match 91.7%; Score 33; DB 2; Length 788;
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKRA 8
 Db 31 AKAARKRA 38

RESULT 4

O8RX24 PRELIMINARY; PRT; 1265 AA.

AC O8RX24: 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical 136.6 kDa protein.
 GN AT4G02475.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eustids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Tang J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Deng C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carriñani P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamlay A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,

RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinzaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Arabidopsis Full Length cDNA Clones";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY090948; AAM13995.1; -
KW Hypothetical protein.
SQ SEQUENCE 1265 AA; 136553 MW; 4272C380DEE70CEF CRC64;

Query Match 91.7%; Score 33; DB 10; Length 1265;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKRA 8
DB 133 SKAARKRA 140

RESULT 5
O9M9F9 PRELIMINARY; PRT; 326 AA.

AC O9M9F9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE F3P9.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;

RN [1]
RP SEQUENCE FROM N.A.
RA Chao O., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
RA Shinn P., Altafi H., Bel Q., Chin C., Chou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharzky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC F3P9 from chromosome
RT I.",
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]

RA Ecker J.R.;
RP SEQUENCE FROM N.A.
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]

RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bel Q., Chin C., Chou J., Choi E.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharzky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC013430; AAF71812.1; -

SQ SEQUENCE 326 AA; 36296 MW; 0CAC70A112745F78 CRC64;

Query Match 88.9%; Score 32; DB 10; Length 326;
Best Local Similarity 87.5%; Pred. No. 86;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKRA 8
DB 96 AEAARKRA 103

RESULT 6
O9EVN5 PRELIMINARY; PRT; 329 AA.

AC O9EVN5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Rnfc protein (Fragment).
GN Rnfc.
OS Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=316;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AL5;
RA Desnoves N., Lin M., Elmerich C.;
RT "Organisation of nif genes in Pseudomonas stutzeri AL5, a rice
RT endophyte."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ297529; CAC03726.1; -

DR HSSP: P00198; 1FCA.
DR InterPro: IPR001450; 4Fe4S_ferredoxin.
DR InterPro: IPR001949; Complex1_51K.
DR Pfam: PF01512; Complex1_51K; 1.
DR Pfam: PF00037; fer4; 2.
DR PROSITE: PS00198; 4Fe4S_FERREDOXIN; 1.
FT NON_TER 1
SQ SEQUENCE 329 AA; 34859 MW; F5888C70A89AC7AE CRC64;

Query Match 88.9%; Score 32; DB 2; Length 329;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKR 7
DB 311 AKAARKR 317

RESULT 7
O8S5V9 PRELIMINARY; PRT; 542 AA.

AC O8S5V9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative serine/threonine kinase.
GN OJ1015F07.8.

OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eurnatoidae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;

RN [1]
RP SEQUENCE FROM N.A.
RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Sasaki C., Henry D., Oates R., Simmons J.;

RT "Rice Genomic Sequence."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC104427; AAM19110.1; -

KW Kinase.
SQ SEQUENCE 542 AA; 61152 MW; 0462D00A2F4427F3 CRC64;

Query Match 88.9%; Score 32; DB 10; Length 542;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARKR 7
DB 450 AKAARKR 456

RESULT 8
O52693 PRELIMINARY; PRT; 60 AA.

AC O52693;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE NifD protein (Fragment).
 GN NIFD.
 OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodobacter
 NC NCBL_TaxID=1061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94040794; PubMed=7693551;
 RA Willison J.C., Pierard J., Huebner P.;
 RT "Sequence and transcript analysis of the nitrogenase structural gene
 RT operon (nifHDK) of Rhodobacter capsulatus: evidence for intramolecular
 RT processing of nifHDK mRNA."
 RL Gene 133:39-46(1993).
 DR EMBL; X63352; CAA44955.1; -
 DR HSSP; P07328; 3MIN.
 FT NON_TER 60
 SQ SEQUENCE 60 AA; 6177 MW; AB97EE643CB6A3A0 CRC64;

Query Match 86.1%; Score 31; DB 10; Length 60;
 Best Local Similarity 87.5%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
 DB 25 AKAKKRA 32

RESULT 9
 ID 08VJ34 PRELIMINARY; PRT; 821 AA.
 AC 08VJ34;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE AT425730/Fl4M19.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCBL_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shim P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
 RA Banb J., Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY064128; AAL36036.1; -
 DR InterPro; IPR000637; AT_hook.
 DR InterPro; IPR002877; FtsJ.
 DR InterPro; IPR001092; HLH_basic.
 DR Pfam; PF01728; FtsJ; 1.
 DR PRINTS; PR00929; ATHOOK.
 DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 SQ SEQUENCE 821 AA; 92366 MW; 047FA266CD5395E6 CRC64;

Query Match 86.1%; Score 31; DB 10; Length 821;
 Best Local Similarity 87.5%; Pred. No. 3.3e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
 DB 705 AKAKKRA 712

RESULT 10

O9SW10
 ID O9SW10 PRELIMINARY; PRT; 821 AA.
 AC O9SW10;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 92.2 kDa protein (Fragment).
 GN Fl4M19.10 OR AT4G25730.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCBL_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,
 RA Hohenseil J., Mewes H.W., Mayer K.F.X., Schueller C.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Pohl T., Weizenegger T., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Mewes H.W.,
 RA Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL049480; CAB39594.1; -
 DR EMBL; AL161563; CAB81383.1; -
 DR HSSP; P28692; IEJ0.
 DR InterPro; IPR002877; FtsJ.
 DR InterPro; IPR001092; HLH_basic.
 DR Pfam; PF01728; FtsJ; 1.
 DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 KW Hypothetical protein.
 FT NON_TER 821
 SQ SEQUENCE 821 AA; 92186 MW; 182617C276C4AFC CRC64;

Query Match 86.1%; Score 31; DB 10; Length 821;
 Best Local Similarity 87.5%; Pred. No. 3.3e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKAKKRA 8
 DB 705 AKAKKRA 712.

RESULT 11
 ID 08XWK6 PRELIMINARY; PRT; 70 AA.
 AC 08XWK6;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical transmembrane protein RSC2468.
 GN RSC2468 OR RS01137.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 NC NCBL_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,

RA Ariat M., Billault A., Brothier P., Camus J.C., Catolico L.,
 RA Chander M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weisenbach J., Boucher C.A.,
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*,"
 RL Nature 415:497-502(2002).
 DR EMBL: AL646070; CAD16175.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 70 AA; 7718 MW; 70EB3E4FC04857AF CRC64;

Query Match 83.3%; Score 30; DB 16; Length 70;
 Best Local Similarity 75.0%; Pred. No. 53;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKAARKRA 8
 DB 19 AKASRRRA 26

RESULT 12

OBVWT1 PRELIMINARY; PRT; 106 AA.
 AC Q8VWT1;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE HMG-domain containing protein (Fragment).
 OS Narcissus pseudonarcissus (Dafodil).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
 OC Narcissus.
 ON NCBI_TaxID=39639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. DUTCH MASTER; TISSUE=TEPAL;
 RA Hunter D.A., Reid M.S.;
 RT "Identification of genes associated with perianth senescence in
 RT daffodil (*Narcissus pseudonarcissus* L. 'Dutch Master').";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF62216; AAL69379.1; -;
 DR InterPro: IPR000910; HMG_12_box.
 DR Pfam: PF00505; HMG_box; 1.
 DR SMART: SM00398; HMG; 1.
 DR NON_TER 1
 FT INTERPRO: IPR001440; TPR.
 SQ SEQUENCE 106 AA; 11793 MW; E73EF1872BBEB64C CRC64;

Query Match 83.3%; Score 30; DB 10; Length 106;
 Best Local Similarity 75.0%; Pred. No. 78;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKAARKRA 8
 DB 98 AKAKRRA 105

RESULT 13

Q9H8H4 PRELIMINARY; PRT; 198 AA.
 AC Q9H8H4;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE CDNA FLJ13629 fls, clone PLACE1011056, weakly similar to histone H1,
 DE gonadal.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nebekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.,
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK023691; BAB14642.1; -;
 DR InterPro: IPR000637; AT_hook.
 DR SMART: SM00384; AT_hook; 1.
 SQ SEQUENCE 198 AA; 21353 MW; 46F43602BC12487B CRC64;

Query Match 83.3%; Score 30; DB 4; Length 198;
 Best Local Similarity 87.5%; Pred. No. 1,4e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKAARKRA 8
 DB 35 AKAKRRA 42

RESULT 14

Q95SX0 PRELIMINARY; PRT; 208 AA.
 AC Q95SX0;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE LP10092P.
 GN CG5434.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Striplin M., Brokstein P., Hong L., Aghayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno C., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY060452; AAL25491.1; -;
 DR FlyBase: FBgn0038810; CG5434.
 DR FlyBase: FBgn0038810; CG5434.
 DR InterPro: IPR001440; TPR.
 SQ SEQUENCE 208 AA; 23063 MW; 33A445DC25A25541 CRC64;

Query Match 83.3%; Score 30; DB 5; Length 208;
 Best Local Similarity 87.5%; Pred. No. 1.5e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKAARKRA 8
 DB 82 AKAKRRA 89

RESULT 15

Q20220 PRELIMINARY; PRT; 219 AA.
 AC Q20220;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical 24.4 kDa protein.
 GN F40F4.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode *C. elegans*: a platform for
investigating biology. The *C. elegans* Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wilson R.;
RT "The sequence of *C. elegans* cosmid F40F4.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: U40420; AAK84526.2; -
DR InterPro: IPR000182; GCN5acetyltransf.
DR InterPro: IPR003975; Shal_channel.
DR Pfam: PF00583; Acetyltransf; 1.
DR PRINTS: PR01497; SHALCHANNEL.
KW Hypothetical protein.
SQ SEQUENCE 219 AA; 24418 MW; 38578B7CFA005E86 CRC64;

Query Match 83.3%; Score 30; DB 5; Length 219;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 AKAARRA 8
Db 15 SKSARRA 22

Search completed: April 8, 2003, 11:28:24
Job time : 31.4138 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:11:45 ; Search time 25.3448 Seconds
(without alignments)
31.345 Million cell updates/sec

Title: US-09-496-391-16

Perfect score: 27

Sequence: 1 ARAKKA 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	6	21 AAB08140	Peptide modulating
2	27	100.0	6	21 AAB08156	Peptide modulating
3	27	100.0	367	19 AAW1785	Cyclin E (296lu-39
4	27	100.0	386	19 AAW1786	Cyclin E (MYNNNNN
5	27	100.0	392	18 AAW13831	Human cyclin E del
6	27	100.0	394	14 AAR33891	Human cyclin E. H
7	27	100.0	395	16 AAR60223	Cyclin E. Homo sa
8	27	100.0	395	18 AAW25119	Human cyclin E fro
9	27	100.0	395	19 AAW59659	Amino acid sequenc
10	27	100.0	395	20 AAW89672	Human cyclin E. H

11	27	100.0	395	21 AAY77484	Primate protein se
12	27	100.0	395	22 AAB48308	Human cyclin E pro
13	27	100.0	451	23 AAO20507	Protein of App rel
14	27	100.0	493	20 AAY27292	Glucose transporte
15	27	100.0	493	22 AAB30522	Amino acid sequenc
16	27	100.0	616	19 AAW41784	GST-cyclin E-PEST*
17	27	100.0	884	22 ABB62036	Drosophila melanog
18	27	100.0	910	23 AAY90961	DMS:acceptor oxido
19	24	88.9	8	23 ABB74738	Transcription fact
20	24	88.9	8	23 ABB74741	Transcription fact
21	24	88.9	8	23 ABB74751	Transcription fact
22	24	88.9	8	23 ABB74831	Nuclear protein nu
23	24	88.9	10	22 AAG84355	Arabidopsis thalia
24	24	88.9	15	23 AAO13912	Cysteine-X-Cysteine
25	24	88.9	15	23 AAM52784	CCR5-binding pepti
26	24	88.9	26	17 AAR92422	Lytic peptide used
27	24	88.9	26	17 AAR89979	Synthetic lytic pe
28	24	88.9	29	16 AAR84173	Peptide enhancer o
29	24	88.9	55	20 AAY34069	Histone H1 isoform
30	24	88.9	55	21 AAY57367	Human histone H1.5
31	24	88.9	66	21 AAY82574	Copolymer molecula
32	24	88.9	85	22 ABB16093	Novel human diagno
33	24	88.9	88	21 AAG15527	Arabidopsis thalia
34	24	88.9	97	22 ABB95979	Human testicular a
35	24	88.9	97	22 AAM95282	Human reproductive
36	24	88.9	98	21 AAG41068	Zea mays protein f
37	24	88.9	105	21 AAG41067	Human ORFX protein
38	24	88.9	119	23 ABB01235	Human ORFX protein
39	24	88.9	140	21 AAG15839	Arabidopsis thalia
40	24	88.9	140	21 AAG21271	Arabidopsis thalia
41	24	88.9	140	22 ABB69109	Drosophila melanog
42	24	88.9	141	21 AAG11527	Arabidopsis thalia
43	24	88.9	141	21 AAG17646	Arabidopsis thalia
44	24	88.9	141	21 AAG45699	Arabidopsis thalia
45	24	88.9	141	21 AAG54127	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AAB08140	standard; peptide; 6 AA.
11	
AC	AAB08140:
XX	
DT	04-DEC-2000 (first entry)
XX	
DE	Peptide modulating activity of heparin, and other glycans.
XX	
KW	Glycosaminoglycan; proteoglycan; heparin modulation; anticoagulant;
KW	cell attachment; cell adhesion; vein graft; tumour cell metastasis;
KW	cartilage differentiation; wound healing.
XX	
OS	Synthetic.
FT	Key
FT	Misc-difference 1..6
FT	Location/Qualifiers
FT	note="this peptide may be repeated an
FT	unspecified number of times"
XX	
PD	WO200045831-A1.
XX	
PD	10-AUG-2000.
XX	
PF	02-FEB-2000; 2000WO-US02853.
XX	
PR	02-FEB-1999; 99US-0118276.
XX	
PA	(UYE-) UNIV JEFFERSON THOMAS.
XX	
XX	Human cyclin E, Homo sa
XX	Human cyclin E fro
PI	San Antonio JD, Verrecchio A, Schlick BP;

DR WPI; 2000-543446/49.
 XX
 PT Novel synthetic peptides with high affinity for glycoaminoglycans and
 PT proteoglycans, useful for modulating heparin, promoting cell
 PT attachment, modulating tumour metastasis and modulating wound healing -
 XX
 PS Disclosure: Page 23; 76pp: English.
 XX
 CC The present sequence represents a synthetic peptide which has a high
 CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
 CC in methods for modulating heparin or other glycoaminoglycans with
 CC anticoagulant activity, promoting cell attachment or adhesion to
 CC natural or synthetic surfaces (especially vein grafts), modulating
 CC tumour cell metastasis, modulating cartilage differentiation, targeting
 CC drugs to epithelial cell surfaces (or to other cells expressing
 CC proteoglycans), modulating enzymes that act on glycoaminoglycan
 CC substrates, affinity purification of bioactive sequences of a
 CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
 CC anti-coagulant functions mediated through glycoaminoglycans, and
 CC modulating wound healing. The peptide may also be used for blocking
 CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
 CC increase heparin half-life in circulation.
 CC
 SO Sequence 6 AA:
 Query Match 100.0%; Score 27; DB 21; Length 6;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ARAKKA 6
 Db 1 ARAKKA 6
 RESULT 2
 AAB08156
 ID AAB08156 standard; peptide; 6 AA.
 AC AAB08156;
 DT 04-DEC-2000 (first entry)
 DE Peptide modulating activity of heparin, and other glycans.
 DE
 KW Glycoaminoglycan; proteoglycan; heparin modulation; anticoagulant;
 KW cell attachment; cell adhesion; vein graft; tumour cell metastasis;
 KW cartilage differentiation; wound healing.
 KW
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1..6
 FT /note= "this peptide may be repeated an
 FT unspecified number of times"
 FT
 PN WO200045831-A1.
 XX
 PD 10-AUG-2000.
 XX
 PF 02-FEB-2000; 2000MO-US02853.
 XX
 PR 02-FEB-1999; 99US-0118276.
 XX
 XX (UYJE-) UNIV JEFFERSON THOMAS.
 PA
 PI San Antonio JD, Verrecchio A, Schick BP.
 DR WPI; 2000-543446/49.
 XX
 PT Novel synthetic peptides with high affinity for glycoaminoglycans and
 PT proteoglycans, useful for modulating heparin, promoting cell
 PT attachment, modulating tumour metastasis and modulating wound healing -
 XX

PS Disclosure: Page 24; 76pp: English.
 XX
 CC The present sequence represents a synthetic peptide which has a high
 CC affinity for glycoaminoglycans and proteoglycans. The peptide is useful
 CC in methods for modulating heparin or other glycoaminoglycans with
 CC anticoagulant activity, promoting cell attachment or adhesion to
 CC natural or synthetic surfaces (especially vein grafts), modulating
 CC tumour cell metastasis, modulating cartilage differentiation, targeting
 CC drugs to epithelial cell surfaces (or to other cells expressing
 CC proteoglycans), modulating enzymes that act on glycoaminoglycan
 CC substrates, affinity purification of bioactive sequences of a
 CC glycoaminoglycan, modifying endothelial cell pro-coagulant or
 CC anti-coagulant functions mediated through glycoaminoglycans, and
 CC modulating wound healing. The peptide may also be used for blocking
 CC tissue uptake of heparin or other glycoaminoglycans in a mammal to
 CC increase heparin half-life in circulation.
 CC
 SO Sequence 6 AA:
 Query Match 100.0%; Score 27; DB 21; Length 6;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ARAKKA 6
 Db 1 ARAKKA 6
 RESULT 3
 AAW41785
 ID AAW41785 standard; protein; 367 AA.
 AC AAW41785;
 DT 06-JUL-1998 (first entry)
 DE Cyclin E (29Glu-395Ala) truncated polypeptide.
 DE
 KW Cyclin E; human; cdk2; cyclin dependent kinase; inhibitor;
 KW cell cycle.
 KW
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO9803649-A1.
 XX
 PD 29-JAN-1998.
 XX
 PF 16-JUL-1997; 97WO-US11728.
 XX
 PR 24-JUL-1996; 96US-0022338.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX
 PI Evans DR, Hollingsworth RE, Rank KB, Sharma SK;
 DR WPI; 1998-120774/11.
 DR N-PSDB; AAV13182.
 XX
 PT Nucleic acid polymer forming constructs of Cyclin E - used for
 PT producing active protein kinase complexes, used to identify
 PT inhibitors of kinase activity
 XX
 PS Claim 48; Fig 9; 81pp: English.
 CC This polypeptide comprises a truncated human cyclin E polypeptide
 CC composed of amino acids Glu29-Ala395 of full-length cyclin E. A
 CC DNA molecule (see AAV13183) coding for the truncated cyclin E,
 CC modified to include an N-terminal His6 tag and a C-terminal
 CC streptavidin tag (see AAW41786) was constructed and the novel cyclin
 CC E construct has been expressed in Escherichia coli. The fusion
 CC protein can be immobilised to an affinity matrix and eluted as a
 CC soluble complex in the presence of cdk2. Such a complex can be

CC used to identify inhibitors of kinase activity; kinases are
CC important enzymes in regulating the cell cycle.

XX Sequence 367 AA;

Query Match 100.0%; Score 27; DB 19; Length 367;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||

Db 330 ARAKKA 335

RESULT 4

AAW41786
ID AAW41786 standard; Protein: 386 AA.

XX AAW41786;

DT 06-JUL-1998 (first entry)

XX Cyclin E (MKNHHNHK)-(29E,395A)-(SAWRHPOFGS) polypeptide.

KW Cyclin E; human; cdk2; cyclin dependent kinase; inhibitor;

XX cell cycle.

OS Homo sapiens.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..9

FT /note="histidine tag"

FT Peptide 377..386

FT /note="streptavidin tag"

XX WO9803649-A1.

XX 29-JAN-1998.

PF 16-JUL-1997; 97WO-US11728.

PR 24-JUL-1996; 96US-0022338.

XX (PHAA) PHARMACIA & UPJOHN CO.

PI Evans DB, Hollingsworth RE, Rank KB, Sharma SK;

XX WPI: 1998-120774/11.

DR N-PSDB; AAV13183.

PT Nucleic acid polymer forming constructs of Cyclin E - used for
PT producing active protein kinase complexes, used to identify
PT inhibitors of kinase activity

PS Claim 49; Page 41-43; 81pp; English.

XX This polypeptide is composed of a truncated human cyclin E
CC polypeptide (see AAW41785) comprising amino acids Glu29-Ala395 of
CC full-length cyclin E, with an N-terminal histidine tag to
CC facilitate purification by immobilised metal affinity
CC chromatography and a C-terminal streptavidin tag. A DNA sequence
CC (see AAV13183) coding for the construct was obtained by PCR
CC amplification (see AAV13184-85) and used to express the polypeptide
CC in Escherichia coli host cells. Soluble polypeptide associated
CC with GroEL was obtained. This complex was active in binding to
CC cdk2, resulting in kinase activity. Recombinant cdk2 bound
CC specifically to matrix-bound recombinant cyclin E polypeptide
CC resulting in elution of a soluble complex of the polypeptide
CC with cdk2. This complex was active in a kinase activity assay,
CC and can be used to identify inhibitors of kinase activity; kinases
CC are important enzymes in regulating the cell cycle.

SQ Sequence 386 AA;

Query Match 100.0%; Score 27; DB 19; Length 386;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||

Db 339 ARAKKA 344

RESULT 5

AAW13831
ID AAW13831 standard; Protein: 392 AA.

XX AAW13831;

DT 28-JUN-1997 (first entry)

XX Human cyclin E delta 9 mutant.

KW Cyclin E; cell cycle; cell proliferation; breast cancer; therapy;

XX antisense.

XX Homo sapiens.

XX WO9709341-A1.

PD 13-MAR-1997.

PF 05-SEP-1996; 96WO-US14357.

PR 07-SEP-1995; 95US-0003357.

XX (HEAL-) HEALTH RES INC.

XX Keyomarsi K;

XX WPI: 1997-192835/17.

DR N-PSDB; AAT59923.

PT Truncated forms of human cyclin E that are constitutively active -
PT and related antisense molecules for treatment of breast cancer by
PT inhibiting their activity

PS Claim 18; Page 68-69; 86pp; English.

XX Human truncated cyclin E delta 9 (AAW13831) and cyclin E delta 148,
CC (AAW13832) are constitutively active forms of cyclin E. Their amino
CC acid sequences were deduced from cDNA clones (AAT59923-24) derived
CC from tumour cell line MD-MB-157. The delta 9 variant has an
CC in-frame deletion corresponding to amino acids 23-25 of native
CC cyclin E. The variant forms of cyclin E can give rise to an active
CC cyclin/cdk2 complex that can cause phosphorylation of substrate at
CC altered checkpoints and thus loss of control during progression
CC from G1 to S phases in tumour cells. Inhibition of the cyclin E
CC variants using e.g. antisense sequences or peptide mimetopes can be
CC used as a means of treating breast cancer.

XX Sequence 392 AA;

Query Match 100.0%; Score 27; DB 18; Length 392;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||

Db 355 ARAKKA 360

RESULT 6

AAW33891
ID AAR33891 standard; Protein: 394 AA.

```

XX AC AAR33891;
XX XX 20-JUL-1993 (first entry)
XX DT
XX DE Human cyclin E.
XX KW Cell division cycle; G1 phase; regulation; triple cln deletion;
XX KM cell division kinase; CDK; CDC2 protein family; ss.
XX OS
XX XX Homo sapiens.
XX PN WO9306123-A.
XX PD 01-APR-1993.
XX PF 16-SEP-1992; 92WO-US07866.
XX PR 20-SEP-1991; 91US-0764309.
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PT (UYRQ) UNIV ROCKEFELLER.
XX PI Cross F, Koff AC, Ohtsubo M, Roberts JM;
XX DR WPI: 1993-117466/14.
XX DT N-PSDB; AAQ38943.
XX PT Isolated human cyclin E nucleic acid - used for developing
XX PT methods for shortening or lengthening cell cycle in mammalian
XX PT cells
XX PS Claim 1; Fig 2; 101pp; English.
XX CC Cyclin E, a new human cyclin, was isolated by complementation of a
XX CC triple cln deletion in Saccharomyces cerevisiae. The polypeptide
XX CC binds and activates a cell division kinase, partic. CDC2 and is
XX CC capable of shortening the G1 phase of a eukaryotic cell cycle.
XX CC Cyclin E showed genetic interactions with the cdc28 gene,
XX CC suggesting that it plays a role at "START" (the entry point into
XX CC the cell division cycle) by interacting with CDC28. In yeast
XX CC containing a cdc28 mutation, START could be achieved by interaction
XX CC of the human genes cdc2-HS and the human homologue of Xenopus CDK2
XX CC with cyclin E. Cyclin E produced in E.coli bound and activated the
XX CC CDC2 protein in extracts from human G1 cells and antibodies against
XX CC cyclin E immunoprecipitated a histone H1 kinase from HeLa cells.
XX SQ Sequence 394 AA:

Query Match 100.0%; Score 27; DB 14; Length 394;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 357 ARAKKA 362

```

```

FH Key Location/Qualifiers
FT Binding-site 129..215
FT FT /note="cyclin box (MRAII) for CDC kinase binding"
XX XX
XX PN US5449755-A.
XX PD 12-SEP-1995.
XX PF 16-SEP-1992; 92US-0764309.
XX PR 16-SEP-1992; 92US-0947311.
XX PR 20-SEP-1991; 91US-0764309.
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PT Cross F, Koff AC, Ohtsubo M, Roberts JM;
XX DR WPI: 1995-327733/42.
XX DT N-PSDB; AAQ98495.
XX PT New isolated human cyclin E nucleic acid - used to develop prods.
XX PT for detecting or modulating the rate of cell proliferation
XX PS Claim 1; Fig 2b-c; 60pp; English.
XX CC This sequence represents human cyclin E. The cDNA encoding this sequence
XX CC was isolated by complementation of a triple cln deletion in
XX CC S. cerevisiae. This sequence binds and activates a cell division cycle
XX CC protein kinase (CDC kinase), such as CDC2, CDC28 and CDK2. Cyclin E
XX CC controls cell growth rate by modulating G1 progression and S phase entry
XX CC using this CDC kinase. Cell growth rates can be altered by increasing or
XX CC decreasing the levels of cyclin E present. By detecting cyclin E levels
XX CC in cells (such as tumour cells), information on the rate of growth of the
XX CC cell may be obtained. The cDNA encoding this sequence may be used to
XX CC develop products that can be used to change the rate of cell
XX CC proliferation. These products include antibodies that bind to the cyclin
XX CC E protein and inhibit the binding of the CDC kinase. The antibodies can
XX CC be used for therapeutic and diagnostic applications.
XX SQ Sequence 395 AA:

Query Match 100.0%; Score 27; DB 16; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
DB 358 ARAKKA 363

```


PR 07-JUN-1995; 95US-0485859.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI Cross F, Koff AC, Ohtsubo M, Roberts JM;
XX WPI: 1997-362928/33.
DR N-PSDB; AAT79856.
XX
XX Screening assay for modulators of cyclin E activity - useful in the
PT study of the cell cycle and cell replication and in developing drugs
PT to treat cell replication disorders
XX
XX Example 2; Column 55-58; 60pp; English.
XX
XX AAW25119 shows human cyclin E which was encoded by cDNA clone H04.
CC The sequence was derived from human glioblastoma cell line U118.
CC Cyclin E was used in a method for identifying cyclin E activity
CC modulators. The method comprised: (a) establishing replicate test and
CC control cultures of cells that express cyclin E; (b) adding a test
CC compound to the test culture but not the control culture; (c) measuring
CC the G1 phase of cells in the test and control cultures; and (d)
CC determining that the test compound modulates or alters cyclin E activity
CC in a cell if the G1 phase measured for the test culture is shorter or
CC longer than the G1 phase measured for the control culture.
CC Compounds identified may be used as drugs for altering cell cycle
CC progression (cell replication) or for promoting it.
XX
SQ Sequence 395 AA;

Query Match 100.0%; Score 27; DB 18; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
| | | | |
DB 358 ARAKKA 363

RESULT 9
AAW59659
ID AAW59659 standard; Protein: 395 AA.
XX
XX AAW59659;
XX
DT 12-OCT-1998 (first entry)
XX
XX Amino acid sequence of the human cyclin E.
DE Human; cyclin E; cell proliferation; antibody; cell division cycle;
KW inhibition; tumour.
XX
OS Homo sapiens.
XX
XX US5783661-A.
XX
XX 21-JUL-1998.
XX
XX 07-JUN-1995; 95US-0522166.
XX
XX 16-SEP-1992; 92US-0947311.
XX 20-SEP-1991; 91US-0764309.
XX 07-JUN-1995; 95US-0522166.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
XX Cross F, Koff AC, Ohtsubo M, Roberts JM;
XX
XX WPI: 1998-427095/36.
XX N-PSDB; AAW41547.
XX
XX Nucleic acids encoding human cyclin E polypeptide(s) - useful for
PT increasing or decreasing cell proliferation and raising antibodies

PT against the enzyme
XX
XX Disclosure; Fig 2A-2C; 63pp; English.
XX
XX This is the amino acid sequence of the human cyclin E protein, used
CC in the method of the invention for increasing or decreasing cell
CC proliferation. Cyclins are enzymes responsible for regulating cell
CC division cycle (CDC), specifically transitions from G1 to S and G2 to
CC M phase. Nucleotides encoding cyclin E are useful for increasing cell
CC growth during proliferation deficiencies and conversely antisense
CC constructs can be used to inhibit the enzymes production during
CC uncontrolled growth e.g. tumours. Polypeptides encoded by the nucleic
CC acids can be used to raise antibodies which can bind specific parts of
CC cyclin E, useful in detection of the enzyme. As the nucleic acids
CC encode different portions of cyclin E, the enzyme can be characterised
CC as to which parts have which activities.
XX
SQ Sequence 395 AA;

Query Match 100.0%; Score 27; DB 19; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
| | | | |
DB 358 ARAKKA 363

RESULT 10
AAW89672
ID AAW89672 standard; Protein: 395 AA.
XX
XX AAW89672;
XX
XX 26-MAR-1999 (first entry)
XX
XX Human cyclin E.
DE Human; cyclin E; cell division kinase complex; detection; tumour cell;
KW immunological binding partner-cyclin E complex; growth rate;
KW immunologic binding partner.
XX
OS Homo sapiens.
XX
XX US5861259-A.
XX
XX 19-JAN-1999.
XX
XX 07-JUN-1995; 95US-0480912.
XX
XX 16-SEP-1992; 92US-0947311.
XX 20-SEP-1991; 91US-0764309.
XX 07-JUN-1995; 95US-0480912.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
XX Cross F, Koff AC, Ohtsubo M, Roberts JM;
XX
XX WPI: 1999-130386/11.
XX N-PSDB; AAX00217.
XX
XX Immunoassay method for determining presence of human cyclin E in
PT biological fluid - involves detecting presence of immunological
PT binding partner-cyclin E complex which is formed by binding of
PT immunologic binding partner with human cyclin E
XX
XX Example 2; Fig 2; 63pp; English.
XX
XX An assay has been developed for determining the presence of human
CC cyclin E in a biological material. The assay comprises: (1) contacting
CC an immunologic binding partner which specifically binds to human
CC cyclin E with the biological material under conditions sufficient to
CC form an immunological binding partner-cyclin E complex; (2) separating

CC unbound immunologic binding partner from the complex; and (3) detecting
CC the presence of the complex to determine the presence of human cyclin E
CC in the biological material. The assay is used for determining the
CC presence of human cyclin E in biological fluid. Detection of the levels
CC of cyclin E in cells such as tumour cells, provides information on their
CC rate of growth. The present sequence represents human cyclin E.
CC
XX
SQ Sequence 395 AA;
Query Match 100.0%; Score 27; DB 20; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ARAKKA 6
Db 358 ARAKKA 363
RESULT 11
AA77484
ID AA77484 standard; Protein; 395 AA.
AC AA77484;
XX
DT 05-JUN-2000 (first entry)
DE Primate protein sequence, SEQ ID NO: 56.
XX
KW Immune disorder; inflammation; allergy; immunosuppressant;
KW antiairrhritic; antirheumatoid; antiinflammatory; dermatological;
KW antithyroid.
XX
OS Primates.
XX
PN WO200001817-A2.
XX
PD 13-JAN-2000.
XX
PE 06-JUL-1999; 99WO-US12366.
XX
PR 06-JUL-1998; 98US-0110938.
PR 13-JUL-1998; 98US-0114466.
PR 23-JUL-1998; 98US-0093897.
PR 12-AUG-1998; 98US-0132968.
PR 18-AUG-1998; 98US-0136214.
PR 11-SEP-1998; 98US-0099999.
XX
PA (SCHE) SCHERING CORP.
XX
PI Bates EM, Lebecque SJF, Murphy EE, Mattson JD, Gorman DM,
PI Hedrick JA, Wang L, Zlotnik A, Murgolo NJ, Greene JR, Johnston JA;
PI Bazan JF, Mahony D, Lees EM;
XX
DR WPI: 2000-171015/15.
DR N-PSDB: AAC92423.
XX
PT New isolated mammalian genes, used to develop products for treating
PT e.g. immune, inflammatory or allergic abnormalities, cancers or
PT degenerative conditions -
XX
PS Disclosure: Page 217-218; 218pp: English.
CC The invention relates to a number of primate and/or rodent proteins, and
CC the genes which encode them. The invention encompasses human dendritic
CC cell prostaglandin transporter (DC-PGT); the TNF (tumour necrosis
CC factor) receptor family-related proteins HDPEA84, HSLD37R and RANKL;
CC human CC chemokine HCC5; human dendritic protein proteins DUBL1 and DUB
CC 12; human MD-1 and human and murine MD-2 proteins, which exhibit the
CC properties of ligands for proteins comprising a leucine-rich motif
CC (LRR); human cyclin E2; CDNs encoding these proteins; and antibodies
CC against these proteins. The proteins can be used for modulating the
CC physiology or development of a cell. They can be used to mediate uptake
CC of substrates (e.g., prostaglandin-like molecules), to modulate or

CC mediate cellular interactions (e.g., induce or prevent trafficking,
CC proliferation, or differentiation of cells), or are intracellular
CC proteins which are important in various cellular processes such as the
CC deubiquitination of proteins or cell cycle regulation. The products can
CC be used for treating medical conditions such as immune, inflammatory or
CC allergic disorders, or abnormal cellular proliferation, for example,
CC cancers or degenerative conditions. They can be used to modulate immune
CC responses in disease states e.g., autoimmune disorders, including
CC rheumatoid arthritis, systemic lupus erythematosus, Hashimoto's
CC autoimmune thyroiditis, as well as acute and chronic inflammatory
CC responses in which T cell activation, expansion, and/or immunological T
CC cell memory play an important role. Sequences AA77463-Y77464,
CC AA77474-Y77475 and AA77484 represent primate proteins of undefined
CC function, AA77462 and AA77481 are rodent proteins of undefined
CC function, and AA77482 is an avian protein of undefined function. These
CC sequences are given in the sequence listing but are not referred to in
CC the remainder of the specification.
CC
XX
SQ Sequence 395 AA;
Query Match 100.0%; Score 27; DB 21; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ARAKKA 6
Db 358 ARAKKA 363
RESULT 12
AAB48308
ID AAB48308 standard; protein; 395 AA.
XX
AC AAB48308;
XX
DT 02-APR-2001 (first entry)
DE Human cyclin E protein.
XX
KW S-phase kinase associated protein; SKP1; SKP2; SKP2-like protein; 2F;
KW CUL-1; cullin; CDC53; p27; cyclin E; Max; Mad; c-Myc; MDM2; p53; Bax;
KW Bad; Bcl-2; tumour; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200075184-A1.
XX
PD 14-DEC-2000.
XX
PE 05-JUN-2000; 2000WO-US15449.
XX
PR 04-JUN-1999; 99US-0137494.
XX
PA (UYVA) UNIV YALE.
XX
PI Zhang H, Tsvetkov LM, Kondo T;
XX
DR WPI: 2001-061703/07.
DR N-PSDB: AAC84620.
XX
PT Modulating polypeptide levels in a cell, diagnosing and treating tumor,
PT involves altering levels of proteins such as S-phase kinase associated
PT proteins 1, 2 and cullin/CDC53 proteins -
XX
PS Claim 5; Page 148-149; 162pp: English.
CC The invention relates to methods of altering the polypeptide levels in a
CC cell, using proteins selected from S-phase kinase associated proteins 1
CC and 2 (SKP1, SKP2), SKP2-like proteins (2F) and CUL-1 (a member of the
CC cullin/CDC53 family of proteins). The method is useful for altering the
CC level of p27, cyclin E, Max, Mad, c-Myc, MDM2, p53, Bax, Bad or Bcl-2
CC polypeptide in a cell. SKP2 and SKP2-like protein levels are useful for
CC detecting tumours, and in monitoring tumor treatment in a mammal. Agents

CC that modulate interactions between SKP and target proteins are useful for
 CC treating tumours.
 XX
 SQ Sequence 395 AA;
 Query Match 100.0%; Score 27; DB 22; Length 395;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ARAKKA 6
 DB 358 ARAKKA 363
 RESULT 13
 AAO20507
 ID AAO20507 standard; Protein: 451 AA.
 XX
 AC AAO20507;
 XX
 DT 27-JUN-2002 (first entry)
 XX
 DE Protein of APP related human homologue hCP38090.
 XX
 KM Neuroprotective; nootropic; transgenic fly; Alzheimer's disease; Abeta;
 KM amyloid precursor protein; tissue-specific expression control; human APP;
 KM APP pathway modulator; gene therapy.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 451 /note="Encoded by T"
 XX
 PM WO200226820-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 01-OCT-2001; 2001WO-EP11345.
 XX
 PR 29-SEP-2000; 2000US-236893P.
 PR 14-JUN-2001; 2001US-298309P.
 XX
 PA (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX
 PI Cohen D, Dengler UJ, Finelli AL, Freuler F, Konsolaki M;
 PI Reinhardt MMH, Zusman S;
 XX
 PS WPI; 2002-315796/35.
 DR N-PSDB; AAK99401.
 XX
 PT New transgenic fly, containing DNA encoding an Abeta portion of human
 PT APP, useful for identifying agents which modulate the APP pathway and
 PT which can be used to treat Alzheimer's disease -
 XX
 PS Example 4; Page 104-105; 129pp; English.
 XX
 CC The invention relates to a transgenic fly whose genome comprises DNA
 CC encoding a polypeptide having the Abeta portion of human amyloid
 CC precursor protein (APP), fused to a signal sequence. The DNA sequence
 CC encodes a 123 (Abeta40) or 129 (Abeta42) amino acid sequence, given in
 CC the specification. The DNA sequence is operably linked to a tissue-
 CC specific expression control sequence. Expression of the sequence gives
 CC the fly an altered phenotype. The purpose of the invention is for
 CC identifying agents that inhibit or promote the expression and/or function
 CC of genes or encoded polypeptides which modify the APP pathway. The agent
 CC is a compound, triple helix DNA, antisense oligonucleotide, double
 CC stranded RNA molecule, ribozyme, or particularly an antibody. It is used
 CC to treat conditions such as Alzheimer's disease. The agent can be used as
 CC an APP pathway modulator or in gene therapy. This sequence represents the
 CC protein of the APP related human homologue hCP38090.
 CC
 XX

SQ Sequence 451 AA;
 Query Match 100.0%; Score 27; DB 23; Length 451;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ARAKKA 6
 DB 378 ARAKKA 383
 RESULT 14
 AAY27292
 ID AAY27292 standard; Protein: 493 AA.
 XX
 AC AAY27292;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE Glucose transporter protein majority sequence.
 XX
 KM Human; glucose transporter; GLUTX; hexose transport; gene therapy;
 KM chromosome mapping; GLUT1.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1.493 /note="residues Xaa are unspecified"
 XX
 PN US5942398-A.
 XX
 PD 24-AUG-1999.
 XX
 PF 26-FEB-1998; 98US-0031392.
 XX
 PR 26-FEB-1998; 98US-0031392.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Tartaglia LA, Weng X;
 XX
 DR WPI; 1999-526192/44.
 XX
 PT A novel mammalian transporter protein and the gene encoding it, are
 PT useful in the diagnosis and treatment of disorders associated with
 PT aberrant sugar transport
 XX
 PS Examples; Fig 3; 48pp; English.
 XX
 CC The invention relates to a human glucose transporter (GLUTX) protein.
 CC The protein can be expressed by standard recombinant methodology. GLUTX
 CC nucleic acids are useful as hybridization probes for detecting the
 CC presence of GLUTX DNA in a sample, useful for diagnosing conditions
 CC associated with aberrant expression levels of GLUTX. The GLUTX gene is
 CC also useful as a therapeutic agent for regulating translation of GLUTX
 CC mRNA, and for treatment of disorders associated with aberrant expression
 CC of GLUTX and aberrant hexose transport. It is useful for generating
 CC GLUTX specific antibodies, identifying agonists and antagonists of GLUTX,
 CC and identifying nucleic acids in other species encoding nucleic acids
 CC homologous to GLUTX. GLUTX nucleic acids are also useful for identifying
 CC the chromosomal location of GLUTX, and as tissue specific markers.
 CC Sequences AAY27287-291 represent different glucose transporter proteins
 CC which were compared with human GLUTX.
 XX
 SQ Sequence 493 AA;
 Query Match 100.0%; Score 27; DB 20; Length 493;
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ARAKKA 6
 DB 378 ARAKKA 383

Db 223 ARAKKA 228

RESULT 15

AAB30522 ID AAB30522 standard; Protein; 493 AA.

XX AAB30522;

DT 06-MAR-2001 (first entry)

DE Amino acid sequence of a consensus GLUT polypeptide.

KM Human; GLUTX; glucose transporter; hexose transport;
tissue-specific marker.

OS Synthetic.

Key Location/Qualifiers

FT Misc-difference 1..493 /note= "Xaa represent any amino acid"

PN US6136547-A.

PD 24-OCT-2000.

PF 26-APR-1999; 99US-0299549.

PR 26-FEB-1998; 98US-0031392.

PA (MILL-) MILLENNIUM PHARM INC.

PI Weng X, Tartaglia LA;

DR WPI; 2001-040236/05.

PT New human glucose transporter protein and nucleic acid encoding the
protein, useful in the diagnosis and treatment of disorders associated
with aberrant hexose transport

PS Disclosure; Columns 63-66; 48pp; English.

CC The present sequence represents a consensus GLUT polypeptide. The
specification describes a human GLUTX polypeptide. GLUTX is a
glucose transporter protein. The GLUTX polypeptides and polynucleotides
are useful in the diagnosis and treatment of disorders associated
with aberrant hexose transport. The GLUTX polypeptide is useful for
determining whether a given disorder is associated with aberrant
expression of GLUTX or activity of GLUTX. The nucleic acid molecule and
the GLUTX polypeptide are useful as diagnostic or therapeutic agents,
or they can be used to generate antibodies or identify small molecules
that, in turn are clinically useful. The GLUTX nucleic acid molecules
are useful for identifying the chromosomal location of GLUTX and as
tissue-specific markers.

SO Sequence 493 AA;

Query Match 100.0%; Score 27; DB 22; Length 493;

Best Local Similarity 100.0%; Pred. No. 4.7e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
111111

Db 223 ARAKKA 228

Search completed: April 8, 2003, 11:23:49
Job time : 27.3448 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:19:38 ; Search time 7.86207 Seconds
(without alignments)
22.454 Million cell updates/sec

Title: US-09-496-391-16

Perfect score: 27

Sequence: 1 ARAKKA 6

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/Backfilltest.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	367	3	US-08-895-707-6
2	27	100.0	386	3	US-08-895-707-7
3	27	100.0	392	1	US-08-706-539-9
4	27	100.0	392	4	US-09-027-007-9
5	27	100.0	395	1	US-08-485-859-2
6	27	100.0	395	1	US-08-706-539-11
7	27	100.0	395	1	US-08-522-166-2
8	27	100.0	395	1	US-08-488-382A-2
9	27	100.0	395	2	US-08-480-912-2
10	27	100.0	395	4	US-09-027-007-11
11	27	100.0	493	2	US-09-031-392-10
12	27	100.0	493	4	US-09-299-549-10
13	27	100.0	493	4	US-09-610-417-10
14	27	100.0	616	3	US-08-895-707-2
15	27	100.0	26	1	US-08-231-730A-46
16	24	88.9	26	2	US-08-505-486-51
17	24	88.9	26	3	US-08-689-489C-46
18	24	88.9	26	3	US-08-801-028-51
19	24	88.9	26	3	US-09-340-154-51
20	24	88.9	26	4	US-09-232-802A-46
21	24	88.9	26	4	US-09-482-611B-51
22	24	88.9	26	5	PCT-US95-04718-46
23	24	88.9	26	5	PCT-US95-09338-51
24	24	88.9	26	5	PCT-US95-09339-51
25	24	88.9	55	3	US-09-041-889-41
26	24	88.9	60	1	US-08-346-849-16
27	24	88.9	60	2	US-08-293-284A-16

28	24	88.9	158	3	US-09-041-889-40	Sequence 40, Appl
29	24	88.9	226	3	US-09-041-889-32	Sequence 32, Appl
30	24	88.9	243	4	US-09-134-001C-4239	Sequence 4239, Ap
31	24	88.9	344	4	US-09-389-341-72	Sequence 72, Appl
32	24	88.9	383	4	US-09-206-800-8	Sequence 8, Appl
33	24	88.9	400	4	US-09-150-347-1	Sequence 1, Appl
34	24	88.9	400	4	US-09-665-313-1	Sequence 1, Appl
35	24	88.9	909	4	US-09-425-383-2	Sequence 2, Appl
36	24	88.9	921	4	US-09-206-800-11	Sequence 11, Appl
37	24	88.9	1150	2	US-08-589-756-3	Sequence 3, Appl
38	24	88.9	1150	4	US-09-206-800-3	Sequence 3, Appl
39	24	88.9	1150	4	US-09-206-898-3	Sequence 2, Appl
40	24	88.9	2396	1	US-08-157-005-2	Sequence 2, Appl
41	24	88.9	2396	4	US-08-747-863-2	Sequence 2, Appl
42	24	88.9	2396	4	US-09-565-864-2	Sequence 2, Appl
43	23	85.2	23	4	US-09-142-355B-16	Sequence 16, Appl
44	23	85.2	23	4	US-09-142-356-4	Sequence 4, Appl
45	23	85.2	58	2	US-08-867-030B-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-895-707-6
Sequence 6, Application US/08895707
Patent No. 6077700
GENERAL INFORMATION:
APPLICANT: (Pharmacia & Upjohn, Co.)
APPLICANT: alternatively, for U.S. filing:
APPLICANT: Hollingsworth, Robert A.
APPLICANT: Sharma, Satish K.
APPLICANT: Rank, Kenneth B.
APPLICANT: Evans, David B.
TITLE OF INVENTION: Special Constructs and Complexes of
NUMBER OF INVENTION: Cyclin E
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESS: Pharmacia & Upjohn Company
STREET: 301 Henrietta Street
CITY: Kalamazoo
STATE: MI
COUNTRY: USA
ZIP: 49001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Rel. #1.0, Ver. #1.25/WordPerfect 5.2+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,707
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Woolton, Thomas A.
REGISTRATION NUMBER: 35,004
REFERENCE/DOCKET NUMBER: 6054
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616-833-7914
TELEFAX: 616-833-8897
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 367 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-895-707-6
Query Match 100.0%; Score 27; DB 3; Length 367;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
Db 330 ARAKKA 335

RESULT 2
US-08-895-707-7
; Sequence 7, Application US/08895707
; Patent No. 6077700
; GENERAL INFORMATION:
; APPLICANT: (Pharmacia & Upjohn, Co.)
; APPLICANT: alternatively, for U.S. filing:
; APPLICANT: Hollingsworth, Robert A.
; APPLICANT: Sharma, Satish K.
; APPLICANT: Evans, Kenneth B.
; APPLICANT: Rank, David B.
; TITLE OF INVENTION: Special Constructs and Complexes of
; TITLE OF INVENTION: Cyclin E
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: MI
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Rel. #1.0, Ver. #1.25/WordPerfect 5.2+
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,707
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woolton, Thomas A.
; REGISTRATION NUMBER: 35,004
; REFERENCE/DOCKET NUMBER: 6054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-7914
; TELEFAX: 616-833-8897
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-895-707-7
Query Match 100.0%; Score 27; DB 3; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ARAKKA 6
|||||
Db 339 ARAKKA 344
RESULT 3
US-08-706-539-9
; Sequence 9, Application US/08706539
; Patent No. 5763219
; GENERAL INFORMATION:
; APPLICANT: Keyomarsi, Khandan
; TITLE OF INVENTION: CYCLIN E VARIANTS AND USE THEREOF

NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,539
FILING DATE: 05-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/003,357
FILING DATE: 07-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Braman, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20894/121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-706-539-9
Query Match 100.0%; Score 27; DB 1; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ARAKKA 6
|||||
Db 355 ARAKKA 360
RESULT 4
US-09-027-007-9
; Sequence 9, Application US/09027007
; Patent No. 6218515
; GENERAL INFORMATION:
; APPLICANT: Keyomarsi, Khandan
; TITLE OF INVENTION: CYCLIN E VARIANTS AND USE THEREOF
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jaecke, Fleischmann & Mugel, LLP
; STREET: 39 State Street
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14614-1310
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/027,007
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/706,539
FILING DATE: 05-SEP-1996
ATTORNEY/AGENT INFORMATION:

NAME: Braman, Susan J
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 87681.98R062
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-262-3640
TELEFAX: 716-262-4133
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-027-007-9

Query Match 100.0%; Score 27; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
|||||
DB 355 ARAKKA 360

RESULT 5
US-08-485-859-2
Sequence 2, Application US/08485859
Patent No. 5645999
GENERAL INFORMATION:
APPLICANT: ROBERTS, JAMES M
APPLICANT: OHTSUBO, MOTOAKI
APPLICANT: KOFF, ANDREW C
APPLICANT: CROSS, FREDERICK
TITLE OF INVENTION: HUMAN CYCLIN E
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON & KINDNESS
STREET: 1420 FIFTH AVENUE, SUITE 2800
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,859
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/947,311
FILING DATE: 16-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/764,309
FILING DATE: 20-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: FHO18598
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206 682 8100
TELEFAX: 206 224 0779
TELEX: 4938023
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: cyclin E amino acid sequence.
US-08-485-859-2

Query Match 100.0%; Score 27; DB 1; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
|||||
DB 358 ARAKKA 363

RESULT 6
US-08-706-539-11
Sequence 11, Application US/08706539
Patent No. 5763219
GENERAL INFORMATION:
APPLICANT: Keyomarsi, Khandan
TITLE OF INVENTION: CYCLIN E VARIANTS AND USE THEREOF
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,539
FILING DATE: 05-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/003,357
FILING DATE: 07-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Braman, Susan J
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20894/121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1636
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 395 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
PUBLICATION INFORMATION:
AUTHORS: Koff, Andrew
AUTHORS: Cross, Fred
AUTHORS: Fisher, Alfred
AUTHORS: Schumacher, Jill
AUTHORS: Legualliec, Katherine
AUTHORS: Philippe, Michel
AUTHORS: Roberts, James M.
TITLE: HUMAN CYCLIN E, A NEW CYCLIN THAT INTERACTS
WITH TWO MEMBERS OF THE CDC2 GENE FAMILY
JOURNAL: CELL
VOLUME: 66
PAGES: 1217-1228
DATE: 20-SEP-1991
US-08-706-539-11

Query Match 100.0%; Score 27; DB 1; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
|||||

DB 358 ARAKKA 363

RESULT 7

US-08-522-166-2
Sequence 2, Application US/08522166

Patent No. 5783661

GENERAL INFORMATION:

APPLICANT: Roberts, J.M.; Ohtsubo, M; Koff, A.C.; Cross, F.

TITLE OF INVENTION: Human Cyclin E

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen, O'Connor, Johnson and Kindness

STREET: 2800 Pacific First Center, 1420 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage

COMPUTER: IBM PC Compatible/Pentium

OPERATING SYSTEM: MS-Windows 3.1

SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/522,166

FILING DATE: June 7, 1995

CLASSIFICATION: 330

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/764,309

FILING DATE: September 20, 1991

ATTORNEY/AGENT INFORMATION:

NAME: Sheiness, Diana K.

REGISTRATION NUMBER: 35,356

REFERENCE/DOCKET NUMBER: FHRO-1-8597

TELECOMMUNICATION INFORMATION:

TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)

TELEFAX: 1-206-224-0779

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 395 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: polypeptide

DESCRIPTION: Cyclin E amino acid sequence; Figure 2

ORIGINAL SOURCE:

ORGANISM:

IMMEDIATE SOURCE:

US-08-522-166-2

Query Match

Best Local Similarity 100.0%; Score 27; DB 1; Length 395;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6

DB 358 ARAKKA 363

RESULT 8

US-08-488-382A-2
Sequence 2, Application US/08488382A

Patent No. 5807698

GENERAL INFORMATION:

APPLICANT: Roberts, J.M.; Ohtsubo, M; Koff, A.C.; Cross, F.

TITLE OF INVENTION: Human Cyclin E

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen, O'Connor, Johnson and Kindness

STREET: 2800 Pacific First Center, 1420 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage

COMPUTER: IBM PC Compatible/Pentium

OPERATING SYSTEM: MS-Windows 3.1

SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,382A

FILING DATE: June 7, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/764,309

FILING DATE: September 20, 1991

ATTORNEY/AGENT INFORMATION:

NAME: Sheiness, Diana K.

REGISTRATION NUMBER: 35,356

REFERENCE/DOCKET NUMBER: FHRO-1-8600

TELECOMMUNICATION INFORMATION:

TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)

TELEFAX: 1-206-224-0779

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 395 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: polypeptide

DESCRIPTION: Cyclin E amino acid sequence; Figure 2

ORIGINAL SOURCE:

ORGANISM:

IMMEDIATE SOURCE:

US-08-488-382A-2

Query Match

Best Local Similarity 100.0%; Score 27; DB 1; Length 395;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6

DB 358 ARAKKA 363

RESULT 9

US-08-480-912-2
Sequence 2, Application US/08480912

Patent No. 5861259

GENERAL INFORMATION:

APPLICANT: Roberts, J.M.; Ohtsubo, M; Koff, A.C.; Cross, F.

TITLE OF INVENTION: Immunoassays for Detection of Human Cyclin E

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen, O'Connor, Johnson and Kindness

STREET: 2800 Pacific First Center, 1420 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage

COMPUTER: IBM PC Compatible/Pentium

OPERATING SYSTEM: MS-Windows 3.1

SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/480,912

FILING DATE: June 7, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/764,309

FILING DATE: September 20, 1991

ATTORNEY/AGENT INFORMATION:

NAME: Sheiness, Diana K.

REGISTRATION NUMBER: 35,356

REFERENCE/DOCKET NUMBER: FHRO-1-8599

TELECOMMUNICATION INFORMATION:

TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)

TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; DESCRIPTION: Cyclin E amino acid sequence; Figure 2
; ORIGINAL SOURCE:
; ORGANISM:
; IMMEDIATE SOURCE:
US-08-480-912-2

Query Match 100.0%; Score 27; DB 2; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
Db 358 ARAKKA 363

RESULT 10
US-09-027-007-11
; Sequence 11, Application US/09027007
; Patent No. 6218515
; GENERAL INFORMATION:
; APPLICANT: Keyomarsi, Khandan
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jaeckle Fleischmann & Mugel, LLP
; STREET: 39 State Street
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14614-1310
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,007
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/706,539
; FILING DATE: 05-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Braman, Susan J
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 87681.96R062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-262-3640
; TELEFAX: 716-262-4133
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 395 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-027-007-11

Query Match 100.0%; Score 27; DB 4; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
Db 358 ARAKKA 363

RESULT 11
US-09-031-392-10
; Sequence 10, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkiesohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-031-392-10

Query Match 100.0%; Score 27; DB 2; Length 493;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
Db 223 ARAKKA 228

RESULT 12
US-09-299-549-10
; Sequence 10, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299,549
FILING DATE: 26-APR-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/031,392
FILING DATE: 26-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Melkielejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-299-549-10

Query Match 100.0%; Score 27; DB 4; Length 493;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
DB 223 ARAKKA 228

RESULT 13
US-09-610-417-10
Sequence 10, Application US/09610417
Patent No. 6346374
GENERAL INFORMATION:
APPLICANT: Tartaglia, Louis A.
MENG, XUN
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
GLOTEX AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/610,417
FILING DATE: 05-Jul-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/299,549
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Melkielejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-610-417-10

Query Match 100.0%; Score 27; DB 4; Length 493;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
DB 223 ARAKKA 228

RESULT 14
US-08-895-707-2
Sequence 2, Application US/08895707
Patent No. 6077700
GENERAL INFORMATION:
APPLICANT: (Pharmacia & Upjohn, Co.)
APPLICANT: alternatively, for U.S. filing:
APPLICANT: Hollingsworth, Robert A.
APPLICANT: Sharma, Satish K.
APPLICANT: Rank, Kenneth B.
APPLICANT: Evans, David B.
TITLE OF INVENTION: Special Constructs and Complexes of
TIME OF INVENTION: Cyclin E
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pharmacia & Upjohn Company
STREET: 301 Henrietta Street
CITY: Kalamazoo
STATE: MI
COUNTRY: USA
ZIP: 49001

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Rel. #1.0, Ver. #1.25/WordPerfect 5.2+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,707
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Woonton, Thomas A.
REGISTRATION NUMBER: 35,004
REFERENCE/DOCKET NUMBER: 6054
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616-833-7914
TELEFAX: 616-833-8897
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-895-707-2

Query Match 100.0%; Score 27; DB 3; Length 616;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
DB 579 ARAKKA 584

RESULT 15
US-08-231-730A-46
Sequence 46, Application US/08231730A

Patent No. 5561107
GENERAL INFORMATION:
APPLICANT: JAYNES, JESSE M.
APPLICANT: JULIAN, GORDON R.
TITLE OF INVENTION: METHOD OF ENHANCING WOUND HEALING BY STIMULATING FIBROBLAST AN
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: STEVEN J. HULTQUIST
ADDRESSEE: INTELLECTUAL PROPERTY/TECHNOLOGY LAW
STREET: 200 PARK DRIVE, SUITE 210
STREET: P.O. BOX 14329
CITY: RESEARCH TRIANGLE PARK
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE
COMPUTER: APPLE MACINTOSH
OPERATING SYSTEM: MACINTOSH
SOFTWARE: M.S. WORD 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/231,730A
FILING DATE: 04-20-94
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,476
FILING DATE: 04-08-94
APPLICATION NUMBER: 08/039,620
FILING DATE: 06-04-93
APPLICATION NUMBER: 08/148,491
FILING DATE: 11-08-93
APPLICATION NUMBER: 08/148,889
FILING DATE: 11-08-93
ATTORNEY/AGENT INFORMATION:
NAME: HULTQUIST, STEVEN J.
REGISTRATION NUMBER: 28021
REFERENCE/DOCKET NUMBER: 4013-106
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)990-9531
TELEFAX: (919)990-9532
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 26
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE:
DESCRIPTION: PEPTIDE
HYPOTHETICAL: NO
FRAGMENT TYPE: COMPLETE PEPTIDE
ORIGINAL SOURCE: SYNTHETIC
IMMEDIATE SOURCE: SYNTHETIC
PUBLICATION INFORMATION: NOT PREVIOUSLY PUBLISHED
US-08-231-730A-46

Query Match 88.9%; Score 24; DB 1; Length 26;
Best Local Similarity 83.3%; Pred. No. 57;
Matches 5; Conservative 1; Mismatches 0; Indels 0;

OY 1 ARAKKA 6
Db 2 ARAKKA 7

Search completed: April 8, 2003, 11:31:21
Job time : 9.86207 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:28:39 ; Search time 8.27586 Seconds
(Without alignments)
44.324 Million cell updates/sec

Title: US-09-496-391-16

Perfect score: 27

Sequence: 1 ARAKKA 6

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2-6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2-6/ptodata/1/pubppaa/PCF_NEW_PUB.pep:*
3: /cgn2-6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2-6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
5: /cgn2-6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2-6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2-6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
8: /cgn2-6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2-6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2-6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2-6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2-6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2-6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2-6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	451	9	US-09-964-899-31
2	27	100.0	493	9	US-09-981-947A-10
3	24	88.9	15	10	US-09-813-653-1
4	24	88.9	66	10	US-09-816-989A-4
5	24	88.9	143	9	US-09-738-626-5675
6	24	88.9	297	10	US-09-789-836-11
7	24	88.9	353	10	US-09-815-242-11328
8	24	88.9	400	12	US-10-055-430-1
9	24	88.9	415	9	US-09-893-519A-41
10	24	88.9	1150	10	US-09-870-122-3
11	23	85.2	46	10	US-09-859-377-4
12	23	85.2	108	9	US-09-832-355A-54
13	23	85.2	126	9	US-10-013-379-36
14	23	85.2	133	10	US-09-731-872-463
15	23	85.2	213	10	US-09-925-300-1718
16	23	85.2	214	10	US-09-844-864-17
17	23	85.2	239	9	US-10-091-504-639
18	23	85.2	239	10	US-09-764-869-639
19	23	85.2	268	9	US-10-027-806-48

20	23	85.2	268	9	US-10-034-623-48	Sequence 48, Appl
21	23	85.2	268	9	US-10-027-801-48	Sequence 48, Appl
22	23	85.2	271	9	US-10-027-806-16	Sequence 16, Appl
23	23	85.2	271	9	US-10-034-623-16	Sequence 16, Appl
24	23	85.2	271	9	US-10-027-801-16	Sequence 16, Appl
25	23	85.2	272	10	US-09-925-297-661	Sequence 661, Appl
26	23	85.2	294	9	US-09-738-626-6320	Sequence 3, Appl
27	23	85.2	294	10	US-09-838-564A-3	Sequence 6320, Appl
28	23	85.2	302	10	US-09-880-192-55	Sequence 55, Appl
29	23	85.2	308	9	US-10-028-072-100	Sequence 100, Appl
30	23	85.2	308	9	US-10-121-049-100	Sequence 100, Appl
31	23	85.2	308	9	US-10-123-904-100	Sequence 100, Appl
32	23	85.2	308	9	US-10-140-470-100	Sequence 100, Appl
33	23	85.2	308	9	US-10-175-746-100	Sequence 100, Appl
34	23	85.2	308	9	US-10-176-918-100	Sequence 100, Appl
35	23	85.2	308	9	US-10-176-921-100	Sequence 100, Appl
36	23	85.2	308	9	US-10-137-865-100	Sequence 100, Appl
37	23	85.2	308	9	US-10-140-474-100	Sequence 100, Appl
38	23	85.2	308	9	US-10-142-431-100	Sequence 100, Appl
39	23	85.2	308	9	US-10-143-114-100	Sequence 100, Appl
40	23	85.2	308	9	US-10-140-002-100	Sequence 100, Appl
41	23	85.2	308	9	US-10-142-419-100	Sequence 100, Appl
42	23	85.2	308	9	US-10-123-262-100	Sequence 100, Appl
43	23	85.2	308	9	US-10-142-423-100	Sequence 100, Appl
44	23	85.2	308	9	US-10-121-050-100	Sequence 100, Appl
45	23	85.2	308	9	US-10-141-755-100	Sequence 100, Appl

ALIGNMENTS

RESULT 1
US-09-964-899-31
Sequence 31, Application US/09964899
Patent No. US2002017446A1
GENERAL INFORMATION:
APPLICANT: Cohen, Dalia et al.
TITLE OF INVENTION: Identification of Genes Involved in
FILE REFERENCE: 4-31612 A
CURRENT APPLICATION NUMBER: US/09/964, 899
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: 60/236, 893
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/298, 309
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 451
TYPE: PRT
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(451)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-964-899-31
Query Match 100.0%, Score 27, DB 9, Length 451;
Best Local Similarity 100.0%, Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ARAKKA 6
Db 378 ARAKKA 383
RESULT 2
US-09-981-947A-10
Sequence 10, Application US/09981947A
Patent No. US20020164578A1
GENERAL INFORMATION:
APPLICANT: Tartaglia, Louis A.

Meang, Xun
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
GLUTEX AND USES THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/981,947A
FILING DATE: 18-Oct-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/031,392
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Melkielejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/072002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-981-947A-10

Query Match 100.0%; Score 27; DB 9; Length 493;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ARAKKA 6
Db 223 ARAKKA 228

RESULT 3
US-09-813-653-1
Sequence 1, Application US/09813653
Patent No. US20020064770A1
GENERAL INFORMATION:
APPLICANT: Nestor, John
APPLICANT: Wilson, Carol
APPLICANT: See, Raymond
APPLICANT: Tan Heht, Christina
TITLE OF INVENTION: Binding Compounds and Methods For Identifying Binding Compounds
FILE REFERENCE: CNS-005
CURRENT APPLICATION NUMBER: US/09/813,653
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/190,946
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/190,996
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/191,299
PRIOR FILING DATE: 2000-03-21
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 15
TYPE: PRT
ORGANISM: Artificial sequence

FEATURE:
OTHER INFORMATION: Small peptide ligand binding compound
US-09-813-653-1

Query Match 88.9%; Score 24; DB 10; Length 15;
Best Local Similarity 83.3%; Pred. No. 16;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ARAKKA 6
Db 10 ARAKKA 15

RESULT 4
US-09-816-989A-4
Sequence 4, Application US/09816989A
Patent No. US20020115103A1
GENERAL INFORMATION:
APPLICANT: Gad, Alexander
APPLICANT: Lis, Doris
TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT M
TITLE OF INVENTION: AND FOR THERAPEUTIC USE
FILE REFERENCE: 2609/60807-A-PCT-US
CURRENT APPLICATION NUMBER: US/09/816,989A
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: 60/101,693
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: PCT/US99/22402
PRIOR FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 66
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-4

Query Match 88.9%; Score 24; DB 10; Length 66;
Best Local Similarity 83.3%; Pred. No. 70;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ARAKKA 6
Db 12 ARAKKA 17

RESULT 5
US-09-738-626-5875
Sequence 5875, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHITAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patentln Ver. 3.0
SEQ ID NO 5875
LENGTH: 143
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5875

Query Match
Best Local Similarity 88.9%; Score 24; DB 9; Length 143;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
Db 50 ARAKKA 55

RESULT 6
US-09-789-836-11
Sequence 11, Application US/09789836
Patent No. US20020082204A1
GENERAL INFORMATION:
APPLICANT: BRIGHAM, KENNETH L.
APPLICANT: STECENKO, ARLENE A.
APPLICANT: SEALY, LINDA
TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20
FILE REFERENCE: N-6977
CURRENT APPLICATION NUMBER: US/09/789, 836
CURRENT FILING DATE: 2001-02-20
PRIOR APPLICATION NUMBER: 60/183,584
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 11
LENGTH: 297
TYPE: PRT
ORGANISM: Rattus sp.
US-09-789-836-11

Query Match
Best Local Similarity 88.9%; Score 24; DB 10; Length 297;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
Db 213 ARAKKA 218

RESULT 7
US-09-815-242-11328
Sequence 11328, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlson, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815, 242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11328
LENGTH: 353
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-815-242-11328

Query Match
Best Local Similarity 88.9%; Score 24; DB 10; Length 353;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
Db 44 ARAKKA 49

RESULT 8
US-10-055-430-1
Sequence 1, Application US/10055430
Patent No. US20020142410A1
GENERAL INFORMATION:
APPLICANT: Rangel-Aldao, Rafael
APPLICANT: Bravo, Adriana
APPLICANT: Sanchez, Beatriz
APPLICANT: Galindo-Castro, Ivan
TITLE OF INVENTION: Malt Beverage Having Stabilized Flavor and Methods of Product
FILE REFERENCE: 1390.0070004
CURRENT APPLICATION NUMBER: US/10/055,430
CURRENT FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: 09/150,347
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/058,398
PRIOR FILING DATE: 1997-09-09
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 1
LENGTH: 400
TYPE: PRT
ORGANISM: Saccharomyces carlsbergensis
US-10-055-430-1

Query Match
Best Local Similarity 88.9%; Score 24; DB 12; Length 400;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
Db 150 ARAKKA 155

RESULT 9
US-09-893-519A-41
Sequence 41, Application US/09893519A
Publication No. US20030027243A1
GENERAL INFORMATION:
APPLICANT: ANADYS PHARMACEUTICALS, INC.
APPLICANT: THOMPSON, Craig
APPLICANT: MOORE, Jeffrey
APPLICANT: BUURMAN, Ed T.
APPLICANT: BRADLEY, John
APPLICANT: DESILVA, Thamara
APPLICANT: HARRIS, Sandra
APPLICANT: KOMARNITSKY, Svetlana
APPLICANT: MENDILLO, Marc

APPLICANT: MOORE, Daniel
APPLICANT: MCCOY, Melissa
APPLICANT: SANDERSON, Karen
APPLICANT: HAO, Tariq
APPLICANT: ZHU, Shuhao
APPLICANT: LONG, Fan
APPLICANT: DAVIDOV, Eugene
TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
FILE REFERENCE: 0342/1G548-US2
CURRENT APPLICATION NUMBER: US/09/893,519A
CURRENT FILING DATE: 2001-06-28
PRIOR APPLICATION NUMBER: US 60/215,164
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 60/224,457
PRIOR FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 146
SOFTWARE: PatentIn version 3.1
SEQ ID NO 41
LENGTH: 415
TYPE: PRT
ORGANISM: Candida albicans
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Corresponds to SEQ ID NO: 114
US-09-893-519A-41

Query Match 88.9%; Score 24; DB 9; Length 415;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
1:||||
Db 161 AKAKKA 166

RESULT 10
US-09-870-122-3
Sequence 3, Application US/09870122
Patent No. US20020142009A1
GENERAL INFORMATION:
APPLICANT: Regents of the University of Minnesota et al.
TITLE OF INVENTION: Streptococcal C5a peptidase vaccine
FILE REFERENCE: 600.450W01
CURRENT APPLICATION NUMBER: US/09/870,122
CURRENT FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: US 09/206,898
PRIOR FILING DATE: 1998-12-07
PRIOR APPLICATION NUMBER: US 08/589,756
PRIOR FILING DATE: 1996-01-22
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 1150
TYPE: PRT
ORGANISM: Streptococcus agalactiae
US-09-870-122-3

Query Match 88.9%; Score 24; DB 10; Length 1150;
Best Local Similarity 83.3%; Pred. No. 1.2e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
1:||||
Db 403 AKAKKA 408

RESULT 11
US-09-859-377-4
Sequence 4, Application US/09859377
Patent No. US20020044946A1
GENERAL INFORMATION:
APPLICANT: TOLLIN AS
APPLICANT: Kisilitchkine, Nikolay

APPLICANT: Jones, Elizabeth L
TITLE OF INVENTION: Product
FILE REFERENCE: 68431/003
CURRENT APPLICATION NUMBER: US/09/859,377
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: RU98120511
PRIOR FILING DATE: 1998-11-18
PRIOR APPLICATION NUMBER: GB9908663.9
PRIOR FILING DATE: 1999-04-15
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 46
TYPE: PRT
ORGANISM: Recombinant Bacteria
FEATURE:
OTHER INFORMATION: Xaa represents any amino acid residue
US-09-859-377-4

Query Match 85.2%; Score 23; DB 10; Length 46;
Best Local Similarity 83.3%; Pred. No. 81;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
||||:1
Db 9 ARAKKA 14

RESULT 12
US-09-832-355A-54
Sequence 54, Application US/09832355A
Publication No. US20030027751A1
GENERAL INFORMATION:
APPLICANT: Kovesdi, Imre
APPLICANT: Kessler, Paul
TITLE OF INVENTION: VEGF FUSION PROTEINS
FILE REFERENCE: 205654
CURRENT APPLICATION NUMBER: US/09/832,355A
CURRENT FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 126
SOFTWARE: PatentIn version 3.0
SEQ ID NO 54
LENGTH: 108
TYPE: PRT
ORGANISM: Homo sapiens
US-09-832-355A-54

Query Match 85.2%; Score 23; DB 9; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RAKKA 6
||||
Db 93 RAKKA 97

RESULT 13
US-10-013-379-36
Sequence 36, Application US/10013379
Publication No. US20020188108A1
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
APPLICANT: Cate, Jamie H.
APPLICANT: No. US20020188108A1ler, Harry F.
APPLICANT: Yusupov, Marat M.
APPLICANT: Yusupova, Guinara ZH
APPLICANT: Baucou, Albion
APPLICANT: Lancaster, Laura
APPLICANT: Dallas, Anne
TITLE OF INVENTION: X-RAY CRYSTAL STRUCTURES OF FUNCTIONAL
TITLE OF INVENTION: COMPLEXES OF THE BACTERIAL RIBOSOME CONTAINING TRANSFER RNA
FILE REFERENCE: 19629-7010


```
; CURRENT APPLICATION NUMBER: US/10/013,379
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: US 60/254,603
; PRIOR FILING DATE: 2000-12-09
; PRIOR APPLICATION NUMBER: US 60/278,013
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/294,394
; PRIOR FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Thermus thermophilus
; FEATURE:
; OTHER INFORMATION: 30S ribosomal protein S13
; OTHER INFORMATION: IgiXP
US-10-013-379-36
```

```
Query Match          85.2%; Score 23; DB 9; Length 126;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ARAKKA 6
    |||||
Db 28 ARAKKA 33
```

```
RESULT 14
US-09-731-872-463
; Sequence 463, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bouguerelet, Lydie
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78-US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 463
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-731-872-463
```

```
Query Match          85.2%; Score 23; DB 10; Length 133;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 RAKKA 6
    |||||
Db 101 RAKKA 105
```

```
RESULT 15
US-09-925-300-1718
; Sequence 1718, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
```

```
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1718
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1718
```

```
Query Match          85.2%; Score 23; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ARAKK 5
    |||||
Db 205 ARAKK 209
```

```
Search completed: April 8, 2003, 11:52:47
Job time : 10.2759 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 8, 2003, 11:18:53 ; Search time 9.51724 Seconds
(without alignments)
60.606 Million cell updates/sec

Title: US-09-496-391-16

Perfect score: 27

Sequence: 1 ARAKKA 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_73:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	161	2 C70613	probable sece - My
2	27	100.0	217	2 E97602	3662 (imported) -
3	27	100.0	217	2 AF2824	transcription regu
4	27	100.0	320	1 A39479	homeotic protein m
5	27	100.0	395	2 A40270	cyclin E - human
6	27	100.0	415	2 A83502	ribonucleoside red
7	27	100.0	424	1 B49851	protochlorophyllid
8	27	100.0	485	2 A84859	probable cytochrom
9	27	100.0	677	2 T39590	hypothetical prote
10	24	88.9	58	2 T13002	ribosomal protein
11	24	88.9	73	2 T16615	hypothetical prote
12	24	88.9	81	2 A11949	DNA topoisomerase
13	24	88.9	81	2 T26945	hypothetical prote
14	24	88.9	97	2 J00527	capsid assembly pr
15	24	88.9	120	2 B90158	conserved hypotet
16	24	88.9	141	2 PS0147	histone H1 - sea u
17	24	88.9	154	2 T04159	histone H1 homolo
18	24	88.9	160	2 C82108	conserved hypotet
19	24	88.9	171	1 HSUR1E	histone H1, gonada
20	24	88.9	172	2 E83301	hypothetical prote
21	24	88.9	180	2 E70565	probable ribosomal
22	24	88.9	206	1 HSTR1R	histone H1 - rainb
23	24	88.9	206	2 S37271	ribosomal protein
24	24	88.9	206	2 S42553	ribosomal protein
25	24	88.9	206	2 S42555	ribosomal protein
26	24	88.9	209	1 HSX1A	histone H1A - Afri
27	24	88.9	217	2 AE0081	3,4-dihydroxy-2-bu
28	24	88.9	218	1 HSCG1	histone H1.02 - ch
29	24	88.9	218	2 S01262	histone H1 - musco

30	24	88.9	219	2 C28456	histone H1.1R - c
31	24	88.9	220	2 A28456	histone H1.10 - ch
32	24	88.9	222	2 T48456	rna binding protei
33	24	88.9	223	2 S49492	histone H1 - mouse
34	24	88.9	224	2 D28456	histone H1.03 - ch
35	24	88.9	225	2 B28456	histone H1.1L - c
36	24	88.9	226	2 S51660	histone H1-5 [vali
37	24	88.9	229	2 S51227	histone H1A - Afri
38	24	88.9	246	2 S26826	histone H1 - maize
39	24	88.9	248	1 HSUR1P	histone H1, gonada
40	24	88.9	250	2 B69802	glucose 1-dehydrog
41	24	88.9	255	2 AE0011	triose-phosphate i
42	24	88.9	289	2 T12682	hypothetical prote
43	24	88.9	297	2 A35914	transcription fact
44	24	88.9	302	2 T07707	hypothetical prote
45	24	88.9	329	2 D71887	ADPglycerolmanno-he

ALIGNMENTS

RESULT 1

C70613 Probable sece - Mycobacterium tuberculosis (strain H37RV)

C.Species: Mycobacterium tuberculosis

C.Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C.Accession: C70613

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A.Reference number: A70500; MUID:98295587; PMID:9634230

A.Accession: C70613

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-161 <COL>

A:Cross-references: GB:292772; GB:AL123456; NID:93261722; PIDN:CAB07097.1; PID:e30657

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: sece

Query Match 100.0%; Score 27; DB 2; Length 161;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ARAKKA 6

Db 86 ARAKKA 91

RESULT 2

E97602 3662 (imported) - Agrobacterium tumefaciens (strain C58, Cereon)

C.Species: Agrobacterium tumefaciens

C.Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002

C.Accession: E97602

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm

A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz,

Science 294, 2323-2328, 2001

A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium

A.Reference number: A97359; PMID:11743194

A.Accession: E97602

A>Status: preliminary

A.Molecule type: DNA

A.Residues: 1-217 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK87774.1; PID:g15157144; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C_3662

A:Map position: circular chromosome

Query Match 100.0%; Score 27; DB 2; Length 217;

Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
Db 190 ARAKKA 195

RESULT 3

transcription regulator, TetR family Atu2020 [imported] - Agrobacterium tumefaciens (str AF2824
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AF2824
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I.
erage, G.; Gillet, M.; Grant, C.; Genthner, D.; Kutyavln, T.; Levy, R.; Li, M.; McCell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF2824
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-217 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL43012.1; PID:g17740475; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu2020
A:Map position: circular chromosome

Query Match 100.0%; Score 27; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
Db 190 ARAKKA 195

RESULT 4
A39479
homeotic protein mec-3 - Caenorhabditis vulgaris
C:Species: Caenorhabditis vulgaris
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 16-Jul-1999
C:Accession: A39479; S24225
R:May, J.C.; Wang, L.; Run, J.Q.; Wang, A.
Genes Dev. 5, 2159-2211, 1991
A:Title: The mec-3 gene contains cis-acting elements mediating positive and negative reg
A:Reference number: A39479; MUID:92084094; PMID:1684166
A:Accession: A39479
A:Molecule type: DNA
A:Residues: 1-320 <MAY>
A:Cross-references: GB:X63956; NID:g11059; PIDN:CAA45377.1; PID:g11060
A:Note: the authors translated the codon CAC for residue 47 as Arg, GCA for residue 171
C:Genetics:
A:Gene: mec-3
A:introns: 26/3; 43/3; 82/3; 106/3; 189/3; 298/3
C:Superfamily: homeotic protein mec-3; homeobox homology; LIM metal-binding repeat homol
C:Keywords: DNA binding; duplication; homeobox; nucleus; transcription regulation; zinc
F:29-79/Domain: LIM metal-binding repeat homology <LIM1>
F:89-145/Domain: LIM metal-binding repeat homology <LIM2>
F:217-273/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 27; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
Db 245 ARAKKA 250

RESULT 5

A40270
cyclin E - human
C:Species: Homo sapiens (man)
C:Date: 17-Jan-1992 #sequence_revision 17-Jan-1992 #text_change 29-Aug-1997
C:Accession: A40270; C40268
R:Koff, A.; Cross, F.; Fisher, A.; Schumacher, J.; Leguellec, K.; Philippe, M.; Robe
Cell 66, 1217-1228, 1991
A:Title: Human cyclin E, a new cyclin that interacts with two members of the CDC2 gen
A:Reference number: A40270; MUID:92005673; PMID:1833068
A:Accession: A40270
A:Molecule type: mRNA
A:Residues: 1-395 <KOF>
A:Cross-references: GB:M73812
R:lew, D.J.; Dulic, V.; Reed, S.I.
Cell 66, 1197-1206, 1991
A:Title: Isolation of three novel human cyclins by rescue of G1 cyclin (Cln) function
A:Reference number: A40268; MUID:92005671; PMID:1833066
A:Accession: C40268
A:Molecule type: mRNA
A:Residues: 1-395 <LEW>
A:Cross-references: GB:M74093
C:Comment: This protein is one of the G1 type cyclins; it forms a complex with both p
C:Genetics:
A:Gene: GDB:CCNE
A:Cross-references: GDB:128967; OMIM:123837
A:Map position: 19q12-19q12
C:Superfamily: cyclin
C:Keywords: cell cycle control

Query Match 100.0%; Score 27; DB 2; Length 395;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
Db 358 ARAKKA 363

RESULT 6
AB3502
ribonucleoside reductase, small chain PA1155 [imported] - Pseudomonas aeruginosa (str
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: AB3502
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; L
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; L
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: AB3502
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-415 <STO>
A:Cross-references: GB:AE004545; GB:AE004091; NID:g9947070; PIDN:AAG04544.1; GSPDB:GN
C:Genetics:
A:Gene: nrDB: PA1155

Query Match 100.0%; Score 27; DB 2; Length 415;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
|||||
Db 55 ARAKKA 60

RESULT 7

B49851

protochlorophyllide reductase (EC 1.3.1.33) 46K chain [similarity] - Rhodobacter capsulatus
 M:Alternate names: chlorin reductase subunit bcHN
 C:Species: Rhodobacter capsulatus
 C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: B49851, S17810
 R:Burke, D.H.; Albert, M.; Hearst, J.E.
 J. Bacteriol. 175, 2414-2422, 1993
 A:Title: bcHFNH bacteriochlorophyll synthesis genes of Rhodobacter capsulatus and Ident
 A:Reference number: A49851; MUID:93224465; PMID:8385667
 A:Accession: B49851
 A:Molecule type: DNA
 A:Residues: 1-424 <BFR>
 A:Cross-references: EMBL:Z11165; NID:946097; PIDN:CAA77526.1; PID:946109
 A:Experimental source: S81003
 A>Note: sequence extracted from NCBI backbone (NCBIT:129238, NCBI:P:129240)
 C:Genetics:
 A:Gene: bcHN
 C:Superfamily: protochlorophyllide reductase chain chLN
 C:Keywords: chlorophyll biosynthesis; oxidoreductase; photosynthesis

Query Match 100.0%; Score 27; DB 1; Length 424;
 Best Local Similarity 100.0%; Pred. No. 98;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
 |||||
 Db 279 ARAKKA 284

RESULT 8
 Probable cytochrome P450 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 23-Mar-2001
 C:Accession: A84859
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKken, S.E.; Umayam, L.; Tallon, L.;
 Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: A84859
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-485 <STO>
 A:Cross-references: GB:A8002093; NID:94512670; PIDN:AAD21724.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g42850
 A:Map position: 2
 C:Superfamily: Synchocytis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: heme; iron; metalloprotein
 F:433/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 100.0%; Score 27; DB 2; Length 485;
 Best Local Similarity 100.0%; Pred. No. 11e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
 |||||
 Db 232 ARAKKA 237

RESULT 9
 T39590
 hypothetical protein SPBC169.16c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T39590
 R:Volckreert, G.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z21865
 A:Accession: T39590
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-677 <VOI>
 A:Cross-references: EMBL:Z99759; PIDN:CA816908.1; GSPDB:GN00067; SPDB:SPBC169.16c
 A:Experimental source: strain 972h; cosmid c1693
 C:Genetics:
 A:Gene: SPDB:SPBC169.16c
 A:Map position: 2
 A:Introns: 1/3; 382/3

Query Match 100.0%; Score 27; DB 2; Length 677;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
 |||||
 Db 145 ARAKKA 150

RESULT 10
 T13002
 ribosomal protein L13, cytosolic - Arabidopsis thaliana
 M:Alternate names: protein T24C20.10
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 29-Oct-1999
 C:Accession: T13002
 R:Choline, N.; Robert, C.; Brotlier, P.; Winkler, P.; Catolico, L.; Artiguenave, F.;
 submitted to the Protein Sequence Database, July 1999
 A:Reference number: Z17586
 A:Accession: T13002
 A:Molecule type: DNA
 A:Residues: 1-58 <CHO>
 A:Cross-references: EMBL:AL096856; GSPDB:GN00061; ATSP:T24C20.10
 A:Experimental source: cultivar Columbia; BAC clone T24C20
 C:Genetics:
 A:Gene: AtSP:T24C20.10
 A:Map position: 3
 C:Keywords: protein biosynthesis; ribosome

Query Match 88.9%; Score 24; DB 2; Length 58;
 Best Local Similarity 83.3%; Pred. No. 90;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6
 |||||
 Db 44 ARAKKA 49

RESULT 11
 T16615
 hypothetical protein K10B3.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C:Accession: T16615
 R:Gallung, S.
 submitted to the EMBL Data Library, February 1996
 A:Description: The sequence of C. elegans cosmid K10B3.
 A:Reference number: Z18546
 A:Accession: T16615
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-73 <GAT>
 A:Cross-references: EMBL:U49941; NID:91206038; PID:91206042; PIDN:AAB53870.1; GSPDB:G
 A:Experimental source: strain Bristol N2; clone K10B3
 C:Genetics:
 A:Gene: CESP:K10B3.1
 A:Map position: X
 A:Introns: 25/3

Query Match 88.9%; Score 24; DB 2; Length 73;
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARAKKA 6

Db 19 ARAKKA 24

RESULT 12

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) precursor - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-May-2000
 C:Accession: A41949
 R:Shiozaki, K.; Yanagida, M.
 Mol. Cell. Biol. 11, 6093-6102, 1991
 A:Title: A functional 125-kDa core polypeptide of fission yeast DNA topoisomerase II.
 A:Reference number: A41949; MID:92049359; PMID:1658625
 A:Accession: A41949
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-81 <SH1>
 A:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hydrolyzing)
 C:Keywords: Isomerase

Query Match 88.9%; Score 24; DB 2; Length 81;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
 Db 37 SRAKKA 42

RESULT 13

hypothetical protein Y45F3A.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T26945
 R:Steward, C.
 Submitted to the EMBL Data Library, October 1998
 A:Reference number: T20289
 A:Accession: T26945
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-81 <WIL1>
 A:Cross-references: EMBL:AL032621; PIDN:CAA21493.1; GSPDB:GN00021; CESP:Y45F3A.5
 A:Experimental source: Clone Y45F3A
 C:Genetics:
 A:Gene: CESP:Y45F3A.5
 A:Map position: 3
 A:Introns: 27/3; 53/3

Query Match 88.9%; Score 24; DB 2; Length 81;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
 Db 42 ARAKKA 47

RESULT 14

J00527
 capsid assembly protein 31-2 - phage T4
 C:Species: phage T4
 A:Note: host Escherichia coli
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 08-Oct-1999
 C:Accession: J00527; J00291; S26172
 R:Filipov, A.G.; Mesyanzhinov, V.V.; Aebi, U.; Kellenberger, E.
 Nucleic Acids Res. 18, 3635, 1990
 A:Title: Cloning and sequencing of bacteriophage T4 genes between map positions 128.3-13
 A:Reference number: J00524; MID:90301484; PMID:2362813
 A:Accession: J00527
 A:Status: translation not shown
 A:Molecule type: DNA

A:Residues: 1-97 <PR1>
 A:Cross-references: EMBL:X17657; MID:g15204; PIDN:CAA35653.1; PID:g15209
 R:Raudonkluene, A.; Nivinskas, R.
 Nucleic Acids Res. 18, 4280, 1990
 A:Title: Nucleotide sequence of bacteriophage T4 gene 31 region.
 A:Reference number: J00290; MID:90332452; PMID:2377483
 A:Accession: J00291
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 26-97 <RAU>
 A:Cross-references: GB:M37882; GB:M23722; MID:g215873; PIDN:AAA32508.1; PID:g215878
 R:Raudonkluene, A.; Nivinskas, R.

Gene 114, 85-90, 1992
 A:Title: Gene rIII is the nearest downstream neighbour of bacteriophage T4 gene 31.
 A:Reference number: S26167; MID:92267389; PMID:1587487
 A:Accession: S26172
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 26-97 <RA2>
 A:Cross-references: EMBL:X54536; MID:g15789; PIDN:CAA38407.1; PID:g15794

Query Match 88.9%; Score 24; DB 2; Length 97;
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
 Db 44 ARAKKA 49

RESULT 15

B90158
 conserved hypothetical protein [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
 C:Accession: B90158
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awaiz, M.J.; Ch
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder
 arett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
 submitted to Genbank, April 2001
 A:Description: Sulfolobus solfataricus complete genome.
 A:Reference number: A99139
 A:Accession: B90158
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-120 <KUR>
 A:Cross-references: GB:AE00641; MID:g13813307; PIDN:AAK40521.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: SSO0175
 C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0429

Query Match 88.9%; Score 24; DB 2; Length 120;
 Best Local Similarity 83.3%; Pred. No. 1.7e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 ARAKKA 6
 Db 69 SRAKKA 74

Search completed: April 8, 2003, 11:30:02
 Job time: 10.5172 secs